Activities Report 2019 of the Research Data Centres (RDCs) accredited by the German Data Forum (RatSWD)

Veröffentlichungsversion / Published Version
Tätigkeitsbericht, Jahresbericht / annual report

This work has been funded by the Federal Ministry of Education and Research of Germany (BMBF) (grant no.: 01UW1802).

Empfohlene Zitierung / Suggested Citation:
Rat für Sozial- und Wirtschaftsdaten (RatSWD). (2020). *Activities Report 2019 of the Research Data Centres (RDCs) accredited by the German Data Forum (RatSWD).* Berlin. https://doi.org/10.17620/02671.59

RatSWD.
German Data Forum

Activities Report 2019

of the Research Data Centres (RDCs) accredited by the German Data Forum (RatSWD)

SPONSORED BY THE

Federal Ministry
of Education
and Research

German Data Forum (RatSWD)

Activities Report 2019

of the Research Data Centres (RDCs) accredited by the German Data Forum (RatSWD)

Contents

Preface by the chairs of the FDI Committee

Making access to research data easier, further developing methods for the collection, documentation, and storage of data, coordinating the research data infrastructure – these are the goals of the network of research data centres (RDCs) that come together in the FDI Committee (`Standing Committee Research Data Infrastructure'). We are glad that our network of RDCs, which consisted of 34 in 2019, expanded by four new RDCs in mid-2020, which has given researchers access to an expanded data offering in the fields of migration, health, financial markets, and, particularly, georeferenced data. This success is built on the work we carried out during the reporting year 2019 that this report is based on. For the future, we want to encourage further institutions to expand the available data for social, behavioural, and economic research by seeking accreditation, and to join the German Data Forum (Rat für Sozial- und Wirtschaftsdaten, RatSWD) network for improving the data infrastructure.

KonsortSWD: Examples of services and measures of and for RDCs relating to data access

- Support for receiving Core Trust Seal certification and for seeking RatSWD accreditation
- Expansion of a network of connected and secure guest researcher workstations
- Development of a federated archiving infrastructure for data from qualitative social research
- Establishment of a training programme for improving data management skills
- Support of a moderated online forum to foster a dialogue between data producers and data users

The FDI Committee has been working on common development goals for several years. These goals have also directly informed the planning of KonsortSWD, the Consortium for the Social, Behavioural, Educational, and Economic Sciences. As part of the large-scale National Research Data Infrastructure project (NFDI), initiated by the Joint Science Conference (GWK), KonsortSWD will, for now, receive funding for five years as of October 2020—a huge success for the RatSWD, the FDI Committee, and others. KonsortSWD will serve as a broad network of universities, research institutes, and, above all, RDCs, and thus create a digital and interconnected knowledge repository. This will, above all, benefit researchers, but also create advantages for RDCs. NFDI funding has enabled us to continue the successful activities of the FDI Committee in cooperation with the RatSWD beyond 2020.

Beyond the FDI Committee, there are now 17 RDCs which have come together in a network that aims at improving the accessibility of research data and quality-assured research data management in the social, behavioural, and economic sciences (*Eco*nomic and *Soc*ial Sciences go*In*g FAIR Implementation Network, EcoSoc-IN). EcoSoc-IN was founded by the RatSWD and is open for national and international research institutions as well as individuals (with or without an accredited research data centre). EcoSoc-IN's goal is to implement the FAIR principles in the social, behavioural, and economic sciences. The FAIR principles aim to make data digitally available by improving the findability, accessibility, interoperability, and re-usability of research data, for humans and particularly for machines. EcoSoc-IN is committed to deepening a dialogue on how to improve access to sensitive data in the social, behavioural, and economic sciences, adjacent disciplines, and on an international level. Its main goal is to standardise access paths and procedures. The network thus supports user-friendly solutions for accessing and linking sensitive data, and, by doing so, addresses the specific demands of the social, behavioural, and economic sciences. In support of RDCs, members of EcoSoc-IN have authored a white paper on the implementation of the FAIR criteria in 2019, which will be published soon.

This year's Activities Report of the RatSWD-accredited research data centres aims to provide a picture of the research data infrastructure in the social, behavioural, and economic sciences. The challenge that the RatSWD-affiliated disciplines routinely face is making data available to researchers that are not originally collected for scientific purposes, and, at the same time, safeguarding the principles of data protection and research ethics. The wide array of methodological approaches, data types, and data topics requires an equally wide array of solutions for documentation, anonymisation, and pseudonymisation. The RDCs' research-based infrastructures set standards for this, therefore making sure that researchers have analysable and well-documented data at their disposal. At the same time, the number of data types is ever increasing, and, with it, the number of specific data curation needs.

With our annual Activities Report, we make sure that our jointly developed standards are being adhered to, thus ensuring a high level of quality, and to give an overview of the activities of RatSWD-accredited RDCs. The report provides answers to important questions: what services do RDCs offer to researchers? How are these services used? How do RDCs make data and metadata (data on datasets) available in the long term and how do they ensure their quality? What is the average number of staff working at RDCs, and what are their tasks? Or quite simply: which RDCs are there, and which data do they provide?

In this year's special topic 'Data linkage', we provide an overview of the extent to which data from different surveys can be or are being joined together, enriching the data foundation for analyses and making new research questions possible. The results show that the data stored in RDCs have great potential for such purposes, but issues of data protection and research ethics must be taken into account.

The RatSWD and the FDI Committee wish to thank the members of the RatSWD's monitoring commission and the team from the RatSWD office for its dedication in responsibly developing this report. By continuously advancing quality assurance processes, the monitoring commission is making an invaluable contribution to better understanding the infrastructure and its capacities, and to making targeted recommendations. The usage figures show that the demand for high-quality data is growing. Modern data infrastructures are increasingly becoming an indispensable part of successful research!

Dr. Pascal Siegers
(Chair of the FDI Committee)

Dr. Jan Goebel
(Co-chair of the FDI Committee)

1 Overview of the research data infrastructure of the German Data Forum (RatSWD)

294
Full-time equivalent staff

521
Scientific publications of RDC staff

4,371
Datasets
431 datasets added

55,270
External Data users

The FDI Committee is a dynamic and decentralised network of 38 research data centres (RDCs) that are accredited by the German Data Forum (RatSWD), four of which were newly accredited in 2020 (as of July 2020). The following report is based on data collected from the 34 RDCs accredited in 2019.[1]

The RDCs archive data and make them available to researchers through several access paths, all the while adhering to data protection regulations. Their spectrum ranges from survey data from various disciplines of the social, behavioural, and economic sciences to geospatial and spatial data, financial data, federal and state-level statistical data, as well as register and social insurance data. In addition to data for quantitative research, qualitative data are now increasingly available too. The network is continuously being expanded to include further RDCs, which thus also increases the range of survey methods, data types, and data formats. The continuous increase in RDCs and data use indicates that this model is fit for the future, and that RatSWD accreditation is seen as a seal of quality by funding agencies and data users, and that it facilitates empirical research.

Current key figures

The key figures of 2019 underscore the successful development of the RDC landscape: on the cut-off date for this publication (31 Dec 2019), the RDCs employed a total of 294 staff in full-time equivalents (FTEs), two thirds of whom were academic staff. RDC staff put out 521 scientific publications in 2019. These are positive numbers in many ways: (re-)using in-house research data creates and secures close ties to the research community, strengthens the support competencies of the RDC staff, thus contributing to quality assurance, and makes the datasets better known. At least 2,359 scientific publications are fully or partly based on the 4,371 datasets made available by the RDCs.

Data users were provided with a total of 431 additional datasets in 2019. While the number of RDCs remained constant, the number of additional datasets per year grew by 62 compared to last year. The potential for the re-use of research data is growing, and data users made good use of this broad array of data in 2019: the RDCs reported 55,270 data users, i.e., 8,600 more than the previous year. Lastly, the RDCs counted 79,878 free downloads—8,400 more than 2018.

Starting point for establishing the research data centre network[2]

The Commission on Improving the Informational Infrastructure (KVI) was established in 1999 as a response to initiatives from within the scientific community.[3] It presented a comprehensive report in March 2001. One of its key recommendations was to set up RDCs at major public data producers, including the federal and state-level statistical offices, the German Pension Insurance, and the Federal Employment Agency, with the aim of professionally archiving existing research data, ensuring they can be used for replicational studies, and responding to new research questions. The Founding Committee leading to the German Data Forum (RatSWD) was set up that same year. This was the corner stone for today's RDC network that the RatSWD has created.

1 See Appendix B for an alphabetical list of the RDCs and more information on the RDCs' data offering (Categories: social, economic, educational, health, psychological, qualitative, and other data).
2 See Appendix A for a chronological overview of the development and services of the research data infrastructure.
3 Zapf, Wolfgang et al. (1996): Memorandum zur Verbesserung der Zugangsmöglichkeiten zu Mikrodaten der amtlichen Statistik. In: ZUMA-Nachrichten, 39, 172–175; Hauser, Richard; Gert G. Wagner and Klaus F. Zimmermann (1998): Erfolgsbedingungen empirischer Wirtschaftsforschung und empirisch gestützter wirtschafts- und sozialpolitischer Beratung. In: Allgemeines Statistisches Archiv 82(3), 369–379.

Evolution and consolidation of the research data infrastructure

Since 2001, the founding of more and more RDCs has led to consistent developments of the research data infrastructure. The reasons for institutions to set up RDCs were multifarious: some sought to implement the recommendations issued by commissions like the KVI, the German Council of Science and Humanities, or scientific advisory groups; other RDCs were commissioned by their parent institutions with the aim of fostering research. What they all had in common was the aim to expand and strengthen the research data infrastructure in Germany by improving access to research data for the scientific community.

The RatSWD was founded as a strategic committee in 2004. To promote a productive dialogue between the RDCs, the RatSWD set up the FDI Committee ('Standing Committee Research Data Infrastructure') in 2009. The main task of this committee is to continuously secure and sustainably improve the research data infrastructure, i.e., expanding the quality and the quantity of data and data access, as well as developing and making data access easier for the research community.

79,878
Downloads
of open access datasets

2,359
Publications
based on RDC data

Info box 1:

Tasks and structures of the German Data Forum (RatSWD)

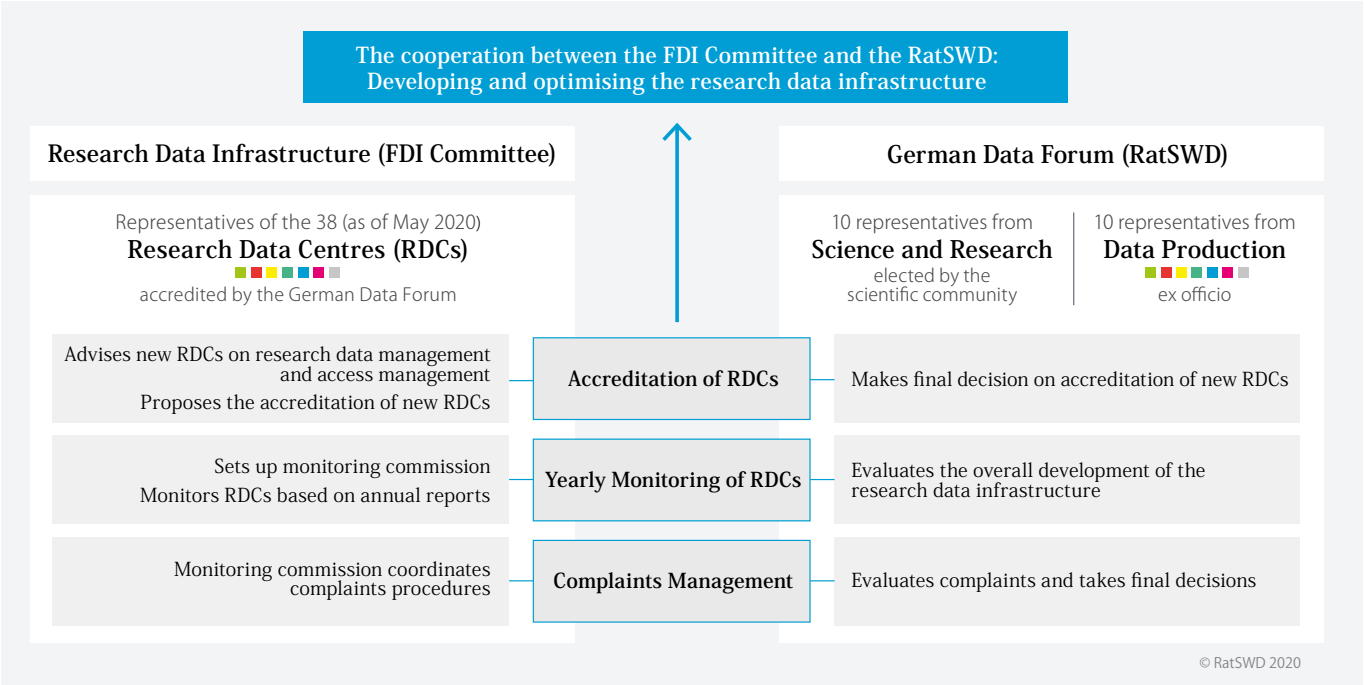
The German Data Forum (Rat für Sozial- und Wirtschaftsdaten, RatSWD) has been advising the German federal government and the governments of the *Länder* in matters concerning the research data infrastructure for the empirical social, behavioural, and economic sciences since 2004. As an institutionalised forum for dialogue, it facilitates a continuous exchange between data producers and data users in science and research with the aim of improving access to high-quality and scientifically potent data. These data are supplied by public, private and scientific actors. From the 7th appointment period and onwards, the RatSWD will consist of ten empirically working scientists who will be appointed based on an election, as well as ten representatives of data-producing institutions of science and administration. This mode of equal representation ensures a broad range of expertise on the committee. The German Data Forum plays a key role in developing research infrastructures in the social, behavioural and economic sciences and is committed to creating research-friendly legal and political frameworks.

At the same time, the RatSWD coordinates 34 research data centres, all of which are accredited according to its guidelines. The research data centres work together in the Standing Committee Research Data Infrastructure, or FDI Committee for short. The RatSWD has created an infrastructure enabling researchers to gain flexible access to a wide range of data.

Accreditation and quality assurance

To ensure the quality of the research data infrastructure, the German Data Forum (RatSWD) defined minimum standards and accreditation criteria in 2010.[4] Since 2015, they have been continuously adjusted to keep up with the technological innovation and the methodological advances within its network. The annual monitoring ensures adherence to standards and high service quality. In addition to annual monitoring, the accreditation applications of new RDCs are supervised by a 'monitoring commission', which is elected by the members of the FDI Committee.[5] The final decision on accreditation is made by the RatSWD. In addition, the RDCs' take part in a complaints management system, which addresses data provision problems that cannot be solved bilaterally; it is administered by the RatSWD office and overseen by the monitoring commission (see Chapter 8).

Fig. 1: The cooperation between the FDI Committee und the RatSWD



4 RatSWD [German Data Forum] (2018): The German Data Forum (RatSWD) and Research Data Infrastructure: Status Quo and Quality Management. RatSWD Output 1 (6). Berlin, German Data Forum (RatSWD). doi: 10.17620/02671.30.

5 For more background information and a list of members of the monitoring commission, see Appendix B.

The results of the monitoring process are compiled in an internal as well as the publicly available Activities Report at hand.[6] German Data Forum (RatSWD) accreditation is a seal of quality for the RDCs. Before accreditation is awarded, RDCs are reviewed for compliance with the RatSWD's mandatory criteria: RDCs must have at least one data access path, datasets must be adequately documented, and they must present a concept for ensuring long-term data availability. Moreover, accreditation also depends on RDCs reporting on tools and materials, quality assurance, further development of the infrastructure, and adherence to data protection regulations.[7] Accreditation strongly benefits the RDCs: they receive extensive support as well as information about best practice solutions to help guide the ongoing development of their own infrastructures, and also participate in the exchange of knowledge and experiences with other RDCs.

Increasing the potential for innovation in research

This close cooperation creates significant potential for innovation in research. Working together in the GWK Joint Science Conference, the federal government and the state governments adopted a recommendation of the German Council for Scientific Information Infrastructures (RfII) to set up the National Research Data Infrastructure (NFDI), a national research data infrastructure, as part of a so-called federal-state accord in 2018. The goal of the NFDI was not primarily to create a technical infrastructure, but to initiate a change in data culture in science and research, and to create a network that views documentation and the sharing of research data as an essential component of good scientific practice. During the 2019 application process, which was coordinated by the German Research Foundation (DFG), the RatSWD successfully put forward a concept for a consortium for the social, behavioural, educational, and economic sciences (KonsortSWD). The RatSWD, FDI Committee, and other key stakeholders of the research data infrastructure worked closely together during this process. In the future, this will lead to a range of key services for RDCs and, especially, the research communities from more than 15 academic associations.

6 All other Activities Reports since 2015 are available at: https://www.ratswd.de/en/publikationen/taetigkeitsberichte

7 Cf. the RatSWD's accreditation criteria RatSWD [German Data Forum] (2017): Qualitätssicherung der vom Rat für Sozial- und Wirtschaftsdaten (RatSWD) akkreditierten Forschungsdatenzentren (FDZ). RatSWD Output 8 (5). Berlin, German Data Forum (RatSWD). DOI: 10.17620/02671.4, 7.

New accreditations

In 2019, the network of 34 accredited RDCs remained stable, and additional RDCs applied for accreditation. Based on the groundwork done in 2019, four new RDCs received preliminary accreditation in the summer of 2020. These new RDCs significantly expanded the existing network and make medical-epidemiological data (GePaRD), financial data (SAFE), spatial data (IOER Monitor), and data on migration and integration issues (DeZIM) available to researchers in the social, behavioural and economic sciences.

German Pharmacoepidemiological Research Database (GePaRD)

The research data centre FDZ GePaRD (German Pharmacoepidemiological Research Database) is a pharmacoepidemiological research database with data from statutory health insurance providers in Germany. Since 2004, the Leibniz Institute for Prevention Research and Epidemiology – BIPS has been working on the establishment and maintenance of GePaRD, which can be used to investigate research questions on the utilization and safety of drugs and vaccines in routine care.

GePaRD contains accounting data from four statutory health insurance providers and information from currently 25 million people, who have been insured through one of them since 2004. In addition to demographic information, GePaRD contains information on reimbursable drug prescriptions, outpatient and inpatient care, and diagnoses. It boasts information on roughly 20% of the general population across all geographical regions in Germany. Ranging from obtaining the data, to preparing them, and making them available, processing takes 25 months, i.e., data from 2018 are ready to use the earliest in late 2020.

BIPS does not own the data and is thus not allowed to decide for which specific projects the data can be used. The approval of projects is based on the authorisation by the health insurance providers and the respective governing authorities. Approval for data use in accordance with § 75 SGB X depends on whether the public interest significantly outweighs the right to personal data protection of the persons concerned. The process of approval by the health insurance providers and the governing authorities usually takes at least three months.

Research Data Center of the Leibniz Institute for Financial Research SAFE

Research on German and European financial markets suffers from a lack of pan-European data sets. Also, existing data sets do not provide a standard identification of, for example, companies. Therefore, researchers often utilize data from the United States where the integration of different databases is more advanced. Consequently, empirical analyses are mostly based on non-European data. However, because of the institutional differences, political recommendations that result from these analyses cannot – or only in a limited scope – be transferred to the European area.

To overcome this problem, the SAFE Research Data Center not only draws on the usual international data sources but also creates new European data sets, combines existing data sets and processes them. The aim is to place the five central research areas of SAFE on a common European data footing.

Monitor of Settlement and Open Space Development (IOER Monitor)

The IOER Monitor is a service of the Leibniz Institute for Ecological Urban and Regional Development (IOER). It provides data and information for land cover and land use structure and change for the whole of Germany, particularly regarding sustainability. The data research centre allows for accessing the data via a browser-based viewer, through geoservices and downloads.

Data is provided on an annual basis and available in high resolution on grids and administrative levels. The data sets provided are based on research results of the IOER. The metadata (data sheet) of the individual data sets give insights on used methods, calculation and used data. Among others, geotopographic data (ATKIS Basis-DLM), land cover data (LBM-DE), official building footprints (HU-DE) and house coordinates (HK-DE) as well as other geospatial data are used as input data.

The Research Data Centre provides access to the data via a map viewer with comprehensive tools and via geo services and downloads. As required the data can be obtained by interested scientists for Germany or for specified spatial delimitations and time periods. An overview of the more than 80 data sets including calculation methods, spatial and temporal reference and the corresponding export functions via geoservices is available at https://www.ioer-monitor.de/en/indicators.

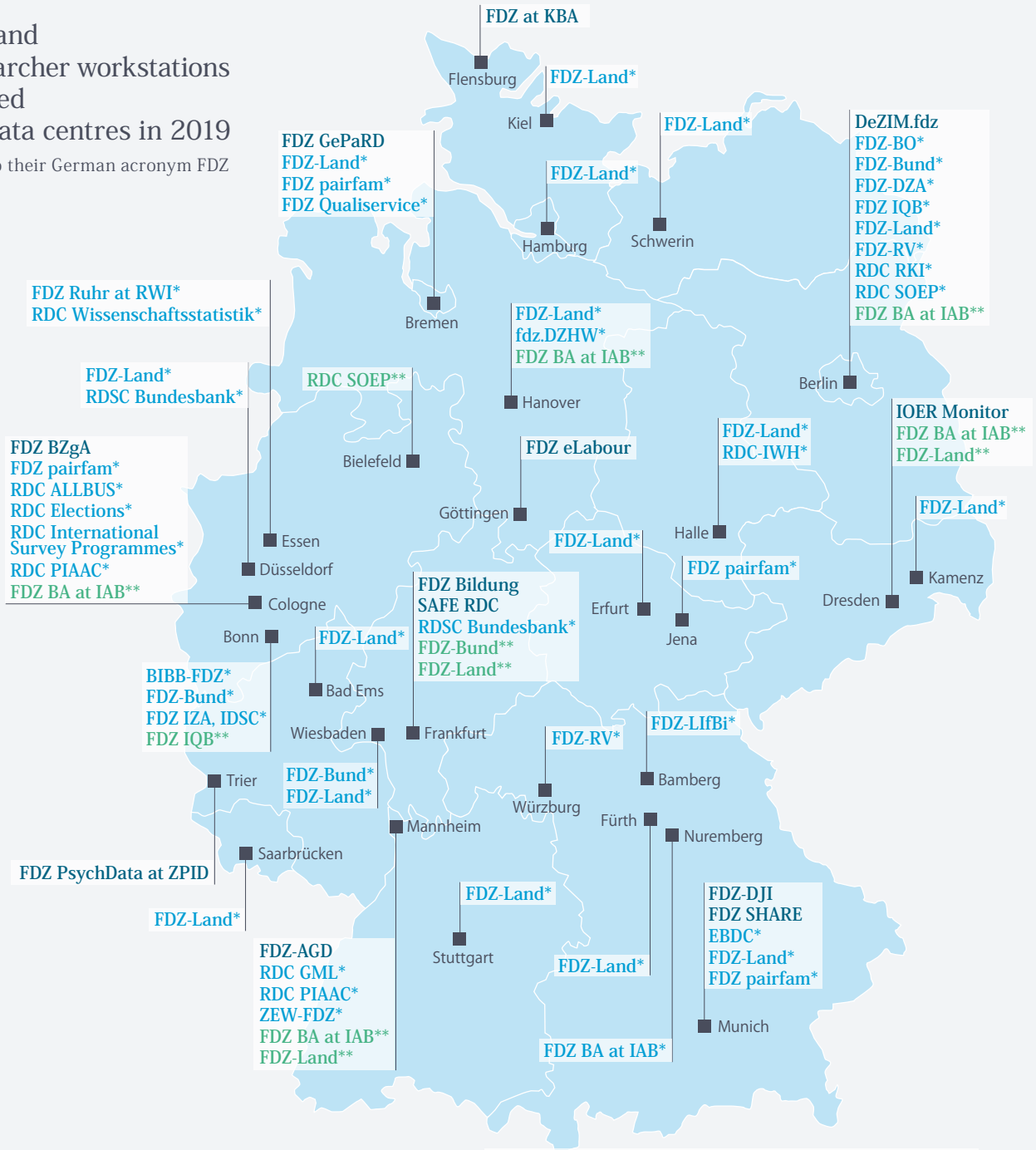
Research Data Centre of the German Centre for Integration and Migration Research (DeZIM)

The German Centre for Integration and Migration Research (DeZIM) is a political and scientific initiative in order to strengthen excellent and internationally visible integration and migration research in Germany. The Research Data Centre DeZIM.fdz gives researchers the opportunity to access data collected within the scope of research projects of the DeZIM institute itself and of the institutes belonging to the DeZIM research community. Besides providing access to these data, the DeZIM.fdz also offers a comprehensive information database. This database allows for research on migration and integration studies archived in the DeZIM.fdz as well as in other research data centres. Moreover, the DeZIM.fdz offers support to data users and gives advice on selected methodological issues.

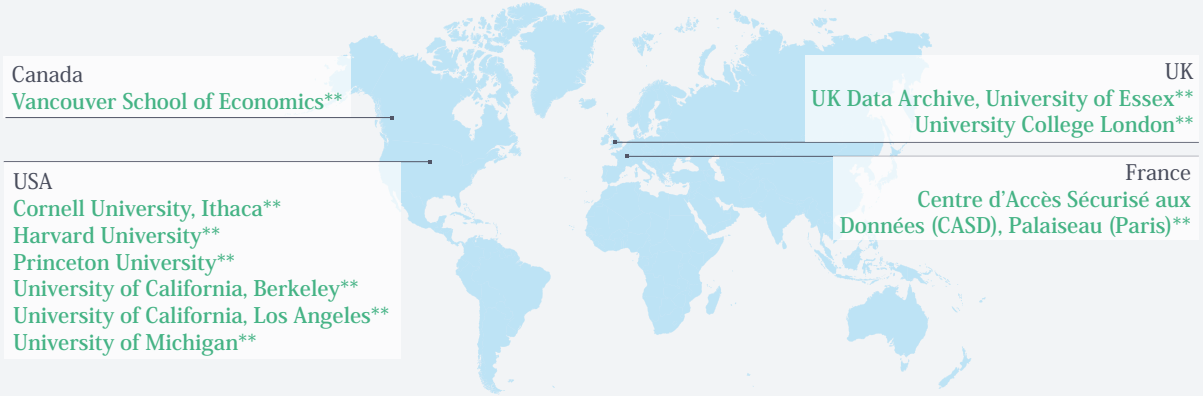
Fig. 2:

Locations and guest researcher workstations of accredited research data centres in 2019

Most RDCs keep their German acronym FDZ



International Locations (FDZ BA at IAB)



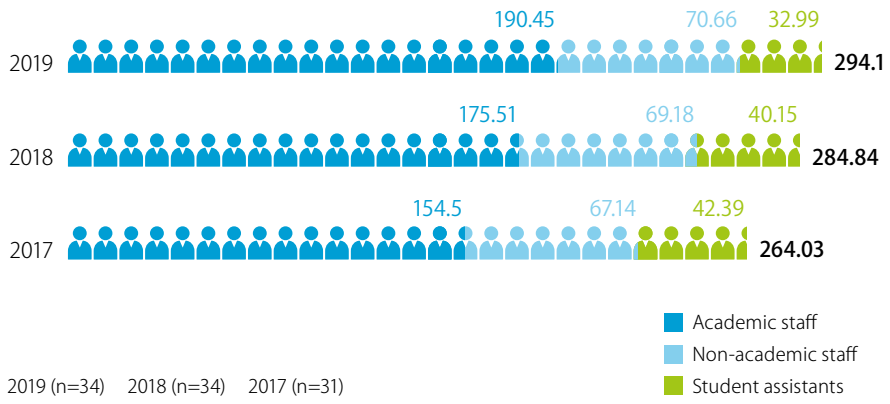
2 Structure of the research data centres (RDCs)

The information presented in the following chapters is gleaned from the RatSWD's annual monitoring process, which was jointly conceived and developed by all RatSWD-accredited research data centres (RDCs), and in which all RDCs participate. In the 2019 reporting year, 34 RDCs took part in the monitoring process.

Staff

The RDCs have been continuously expanding their staff in recent years. As of 31 December 2019, the 34 RDCs employed a total of 294.10 staff in full-time equivalents (FTEs) (see Fig. 3). Much like in previous years, the numbers show that the RDCs have seen an increase in staff, particularly academic staff. The number of student assistants, however, continues to be declining. About a third of RDCs does not employ students at all.

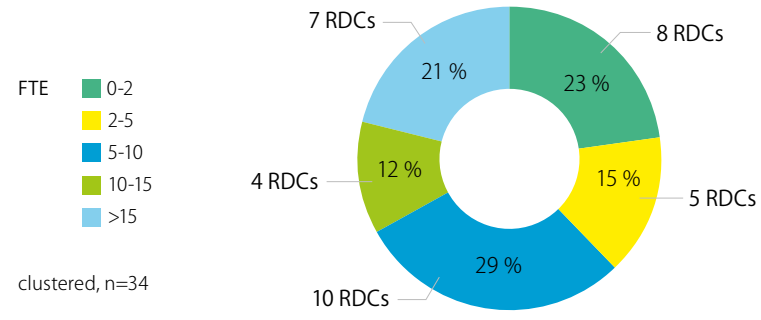
Fig. 3: RDC staff and its distribution in full-time equivalents (FTEs)



There has been a continuous increase in staff over the past three years

The numbers of staff across RDCs vary greatly. They employ an average of 8.9 staff in full-time equivalents, but they can range from less than one FTE to RDCs with 36 FTEs. Just under two thirds of RDCs have ten employees in FTEs at the most, but at least every fifth RDC has more than 15 (see Fig. 4).

Fig. 4: RDC staff in full-time equivalents (FTEs)



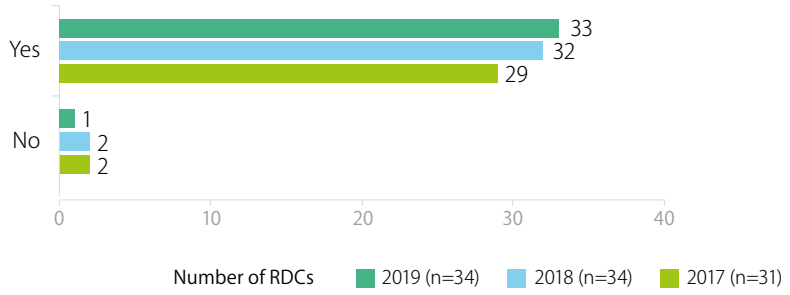
Staff numbers vary strongly across RDCs, but most have five or more FTEs

Since the data in this report is based on the same number of RDCs as in 2018, this increase in staff cannot be attributed to an increase in RDCs. Rather, it can be seen as indicative of changes in data culture, which began some time ago, and have resulted in the readiness of institutions and funding bodies to fund the staff necessary for this change. The rising number of users (see Chapter 4) indeed necessitates this increase. As shown in Fig. 5, this is the case in all but one RDC. These research activities can encompass research-driven issues, the development of methodological innovation, or the advancement of technological developments related to data infrastructure and data services.

Fig. 5: Independent research by academic staff

Does your RDC's academic staff conduct their own research?

At almost every RDC, the academic staff conduct their own independent research

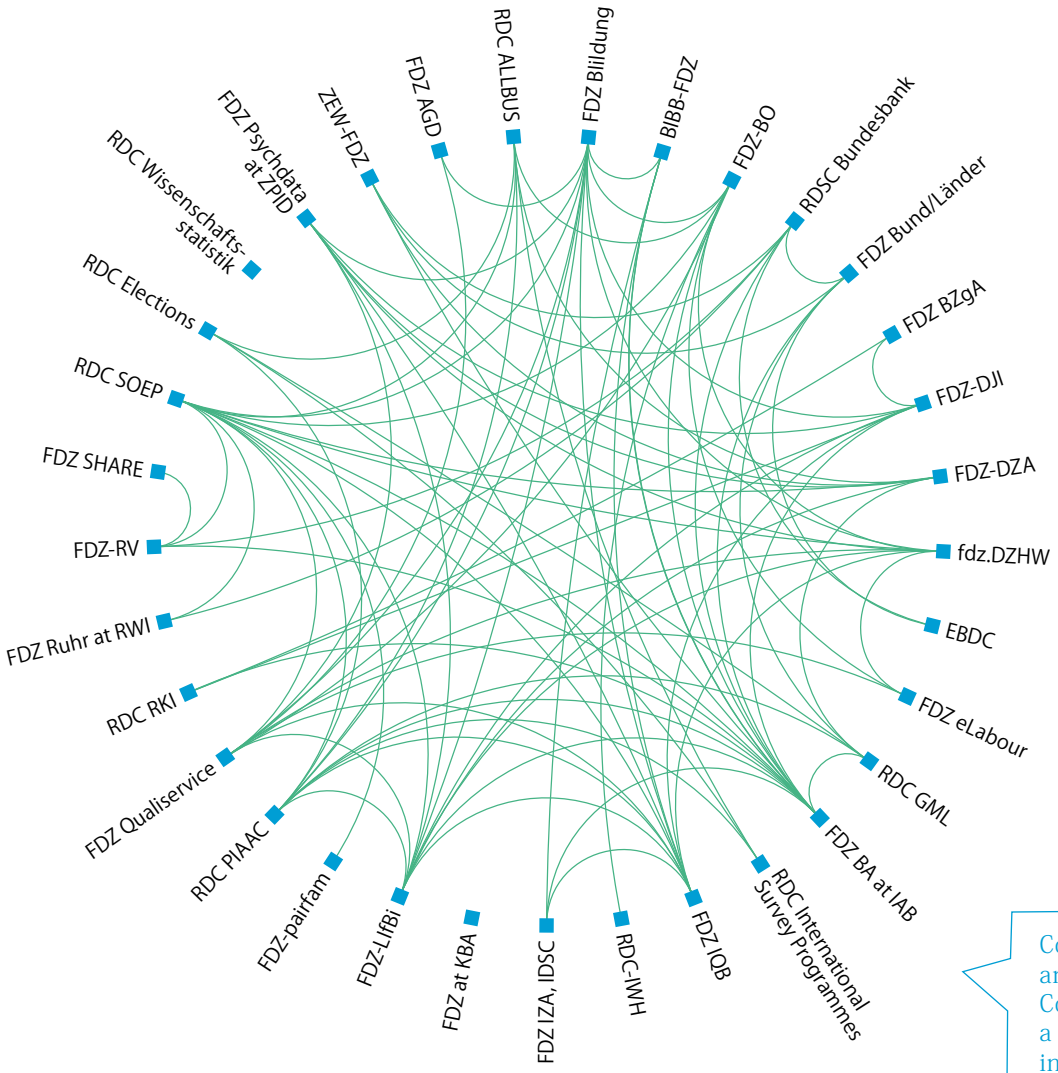


What is the scope of these research activities? At 16 RDCs, the academic staff dedicate a fixed share of their work hours to research activities. On average, this share is just under a third, but there is a broad distribution across the RDCs. At six RDCs, staff dedicate up to one quarter of their work hours to research, and this share is higher at ten RDCs.

Cooperation and research activities

The number of RatSWD-accredited RDCs that maintain institutionalised cooperative relationships with other domestic RDCs has increased compared to the previous year. Twenty-five RDCs reported such relationships, while the number of RDCs that had previously stated they were not involved and did not plan to become involved in such institutionalised relationships fell from 11 to eight RDCs (see Chapter 6, p. 38, for information on international research cooperation). Beyond the activities within the FDI Committee, the RDCs all closely and continuously collaborate with each other (see Fig. 6).

Fig. 6: Cooperation between accredited RDCs in 2019



Cooperation within and beyond the FDI Committee has created a multi-layered, interdisciplinary network of RDCs

Cooperation can either be research-driven and aim to bring together RDCs that focus their data infrastructure on a certain topic (e.g., Verbund FDB, which pools research data for empirical educational research), or are driven by methodological issues. An example of this is VQualidat, a network of RDCs that offer qualitative research data. Cooperation also aims at further implementing the concept of off-site guest researcher workstations (RDC-in-RDC approach). The idea is to allow access from one accredited RDC to data at another accredited RDC to be granted. Lastly, cooperation among RDCs will be deepened by KonsortSWD, the Consortium for the Social, Behavioural, Educational, and Economic Sciences, which has secured funding for participation in the large-scale research data infrastructure project NFDI. The RatSWD's network of RDCs, which all have long-standing experience in operating user-oriented research data infrastructures, will form the backbone of KonsortSWD (see p. 11 for more information on KonsortSWD).

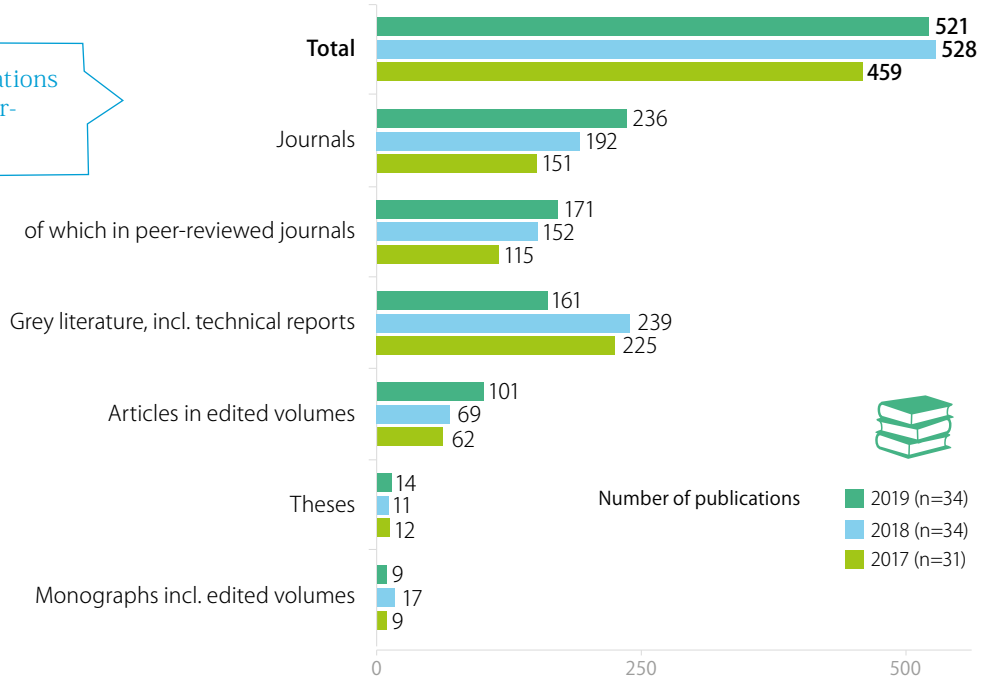
Academic publications

The activities of RDCs are commonly associated with providing data for non-commercial, scientific research. This goes along with providing support services to help users exploit the potential of the available datasets. For this reason, the majority of RDC employees are made up of academic staff, who dedicate a part of their work hours to their own research activities. This enables RDC staff to perform data-related support and consulting services at an adequate scientific level. Research activities are documented by scientific publications. In total, publication output stayed roughly the same compared to the previous year, as shown in Fig. 7, at 521 publications in this reporting year.

Fig. 7: Scientific publications of RDC staff

Please indicate the number of scientific publications produced by your RDC's staff, regardless of the type of data and whether the publication was prepared during RDC working hours.

A third of the publications was published in peer-reviewed journals



The number of publications in journals grew substantially in 2019. In particular, RDC staff were able to publish significantly more articles in peer-reviewed journals, which also indicates the high quality of the research output. The number of theses also grew. Most of these are higher-level theses (PhD, habilitation; not shown separately in the figure). Master's theses also belong in this category. However, since not all RDCs report these, it is safe to assume substantial underreporting here. This is not to say that earlier academic theses, e.g., master's and bachelor's theses, are deemed less important. To the contrary, the Federal Statistical Office has been waiving the already reduced data use costs for theses of young researchers since June 2020. In the entire official statistics community, data access for bachelor's, master's and PhD theses is now free-of-charge.

3 Archiving and quality assurance

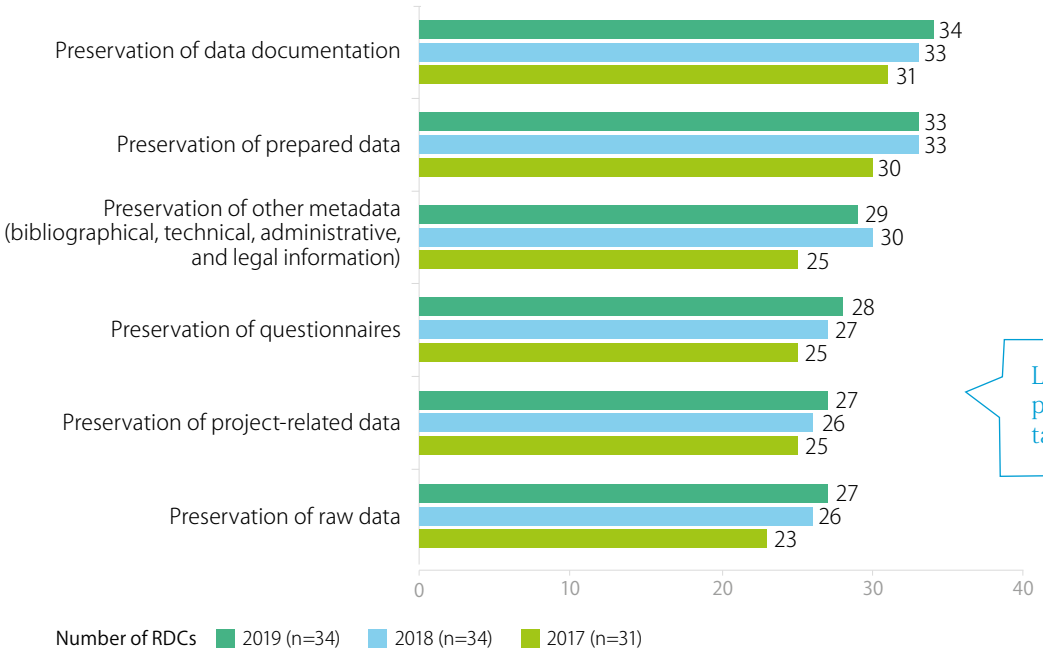
Concepts for making data available on a long-term basis

The long-term availability of data used for research is part of good scientific practice, and therefore firmly anchored in the research institutions. All RDCs thus preserve data documentation and the prepared data (see Fig. 8, the latter is irrelevant to one RDC). Ensuring long-term availability of data after their primary use also contributes to the documentation of research activity, the replicability of empirical research results (for example, in theses or journal articles), and data re-use by third parties.

In addition to physically storing the data (in different versions, if necessary), the RDCs support processes related to data re-use by ensuring that the data can be read, interpreted, and used in the future. Data environments can change over time (e.g., updated versions of the statistics software that was used, where backward computability is not always guaranteed). Therefore, a key challenge is to store data in formats that ensure readability with future technologies. Aside from data archiving, RDCs must also safeguard data against loss as well as unauthorised and unwanted manipulation.

Fig. 8: Data preservation at RDCs

Does your RDC ensure the availability of data on a long-term basis (according to the rules of good scientific practice and for at least ten years after they were last used) and in standard formats? *(Multiple answers possible)*



Long-term data preservation is a core task of the RDCs

Not all RDCs reported preservation of raw data and questionnaires as part of their tasks. This does not mean that these two elements are not preserved, only that these tasks are taken on by others.

Archiving

Thirty-three RDCs use locally redundant storage, i.e., storing data on multiple storage media, and 17 of these RDCs report that they store data in multiple locations. Only one RDC restricts itself to a simple backup on in-house servers. Some empirical studies require updates for data or metadata. The RDCs handle this differently. Most often, RDCs archive and make available all versions (18 RDCs). Eleven RDCs provide users with the most current version of a dataset; five RDCs host datasets that do not require updating.

Ten RDCs have certification for their archiving solution; this number did not change compared to the previous year. Most are certified with the 'Core Trust Seal', a community-based, non-profit organisation. The aim of such certifications is to help researchers identify trustworthy repositories for archiving and curating research data. They provide information on a repository's adherence to transparent standards laid out in the certification criteria. KonsortSWD plans a dedicated work package to foster certification at RDCs and to facilitate the respective processes. The package will focus on certification of RDCs with the Core Trust Seal as well as RatSWD accreditation. The work package therefore aims to offer a service to those RDCs that are not yet part of the RatSWD's network, and to support them in preparing for the accreditation process.

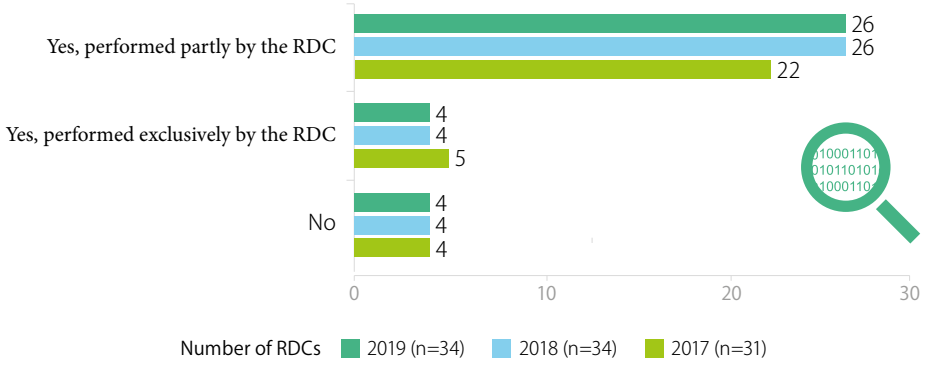
Quality assurance of datasets

In addition to the physical storage of data for use and re-use, one of the staff's key tasks at most RDCs is to assist in data checks. The extent to which RDCs take on these tasks usually depends on whether and how the data centre is affiliated with another (parent) institution. Since these affiliations tend to be stable over time, there have not been any substantial changes to the tasks in this reporting year, as shown in Fig. 9. Much like the previous year, there are four RDCs that are solely in charge of data checks, and about 80% of RDCs perform all quality assurance measures themselves. Four data centres do not perform data checks at all.

The majority of RDCs is involved in data checking

Fig. 9: Data checks at RDCs

Is data checking (checking the quality of shared data) a task of your RDC?



RDCs which are not solely responsible for performing data checking tasks usually share the responsibility of data quality assurance with the primary researchers or the data producers (these can be third-party institutes) (see Fig. 10). Eighteen RDCs have created clear guidelines for data checks.

Fig. 10: Who (or who else) performs data checks?

(Evaluation of an open question, multiple answers possible)

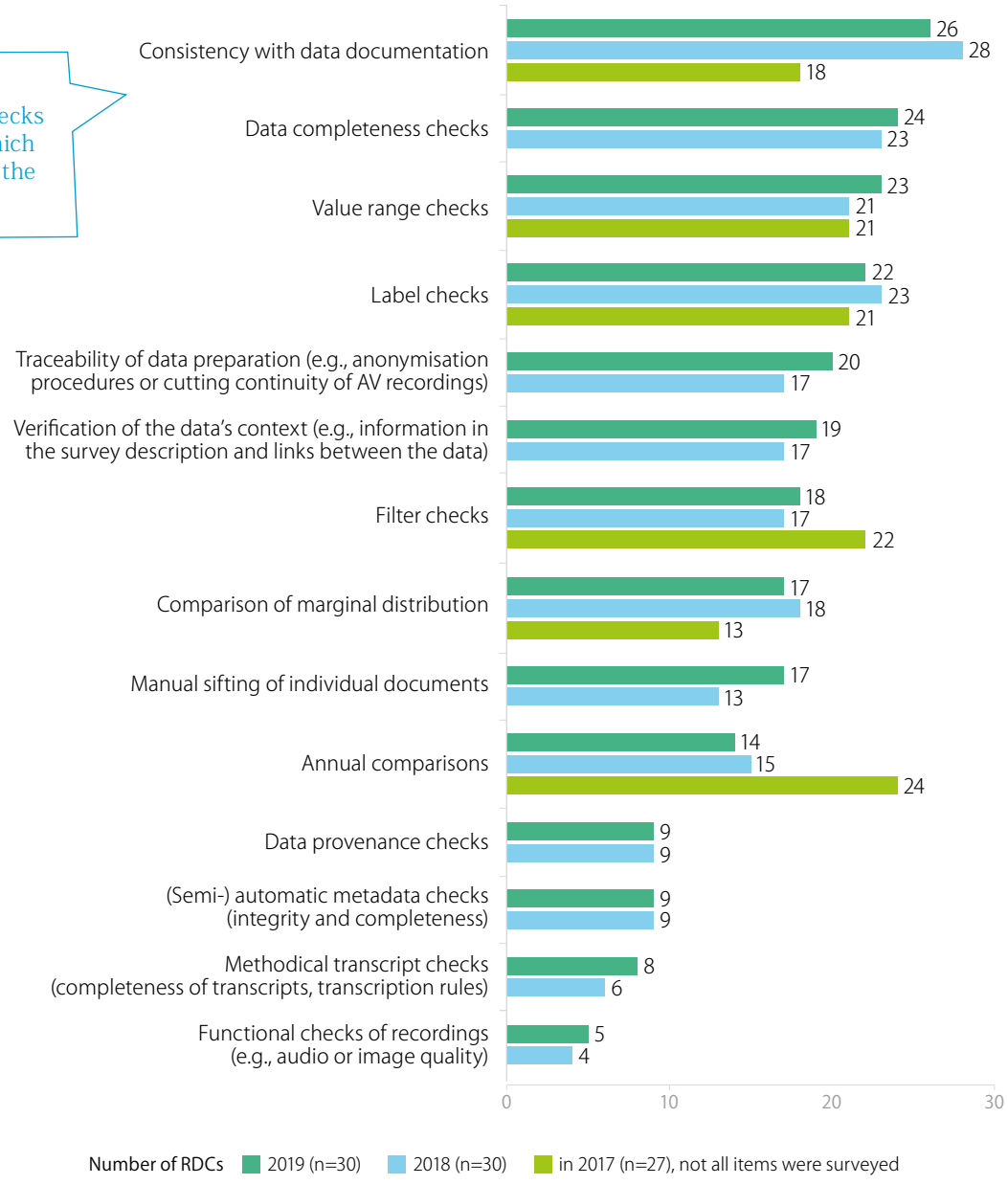
Data checks are performed both internally and externally

The most common data checking measures at RDCs include checking consistency with data documentation and value range checks (see Fig. 11). The number of RDCs reporting that they perform data provenance checks or automatic metadata checks is much lower. One reason might be that data provenance is typically checked before the data are handed over to the RDC. Just because an RDC does not perform certain data checks, does not mean they are not performed at all. Rather, they are often integrated into other stages of the data life cycle. Only few RDCs do methodical transcript checks or functional checks of recordings, because only few accredited RDCs specialise in qualitative empirical research, which is where these techniques are typically used.

The RDCs perform appropriate data checks for each dataset, which strongly depend on the survey method

Fig. 11: Types of data checks

Which types of data checks are performed at your RDC?
(Multiple answers possible)

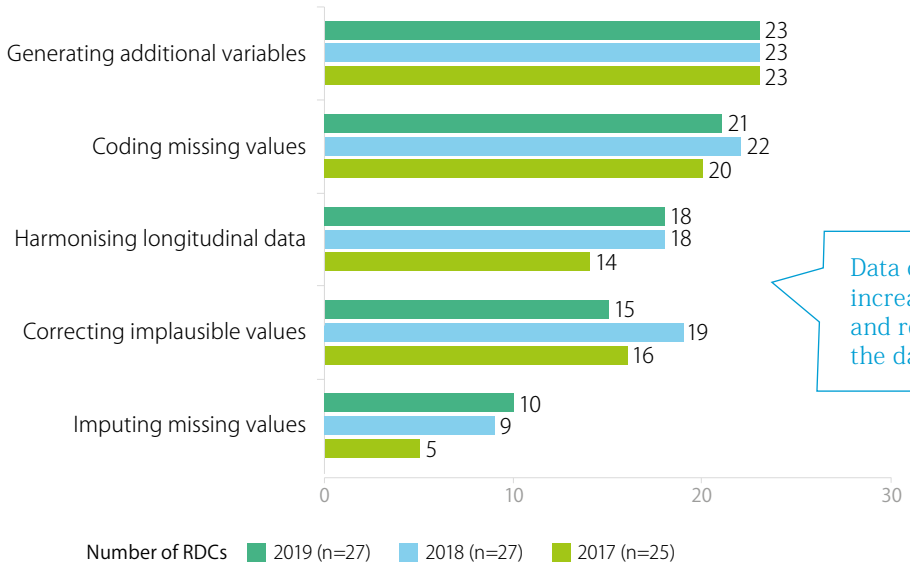


Data checks are an important component of data quality assurance. Another component of data quality assurance is correcting data errors, also during data generation, e.g., to increase user-friendliness. The distribution of responsibilities here is comparable to that for data checks: two RDC correct the data solely themselves; 25 do so only partially. Seven RDCs do not perform data correction measures at all.

At 23 reporting RDCs, the most common task in this area is the generation of additional variables. Highlighting missing values through coding is also widespread (21 RDCs). Harmonisation of longitudinal data is part of the quality assurance measures at 18 RDCs (see Fig. 12). There are clear guidelines for data correction at 15 out of 27 RDC. Data corrections are not always made transparent to users at the individual level, e.g., when checks were performed for data protection or anonymisation reasons.

Fig. 12: Types of data correction at RDCs

Which types of data generation or data correction measures are performed at your RDC?
(Multiple answers possible)



Data correction increase the quality and re-usability of the data

4 Availability and use of data

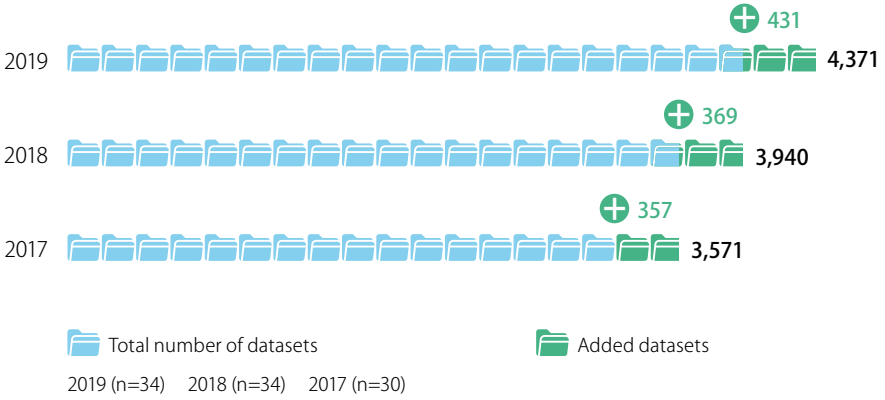
Range of available data

The trove of available data continues to grow: on the cut-off date for this publication (31 December 2019), RDCs made 4,371 datasets available. In the 2019 reporting year, the RDCs added 431 datasets, which were either assigned a new digital object identifier (DOI) or were at least suitable for DOI registration (see Fig. 13). The number of newly added datasets varies across RDCs with some adding datasets in the single digits, and one RDC adding 151 new datasets. Since a dataset can contain several individual studies, the number of available studies is significantly higher.

Fig. 13: Datasets made available by the RDCs

Please indicate the number of new surveys and datasets that were added during the reporting year.
Please note, datasets are regarded as new, for example, if they were assigned a DOI or were suitable for DOI registration.

There has been a continuous increase in available datasets in the past years



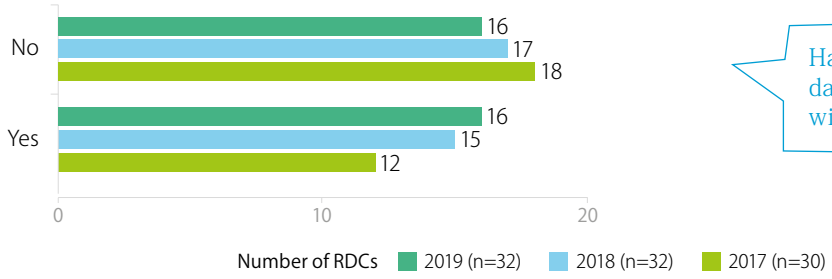
Thirty-one RDCs assign persistent identifiers (PIDs) to their datasets to ensure their long-term findability and citability. All RDCs not yet using PIDs, like DOIs, are currently planning to implement this practice.

Closure periods and fees

A central goal of RDCs is to facilitate low-threshold and timely access to data. Safeguarding equal opportunities, research data should be made available to all qualified users at the same time. High fees and long closure periods for datasets stand in the way of this goal. However, good reasons exist for imposing a closure period on transmitting certain data. Most RDCs argue that closure periods ensure that primary researchers retain the opportunity and right to be the first to utilise the data. Furthermore, closure periods may also protect theses currently being written.

Fig. 14: Closure periods for datasets at RDCs

Do you have closure periods for some datasets?

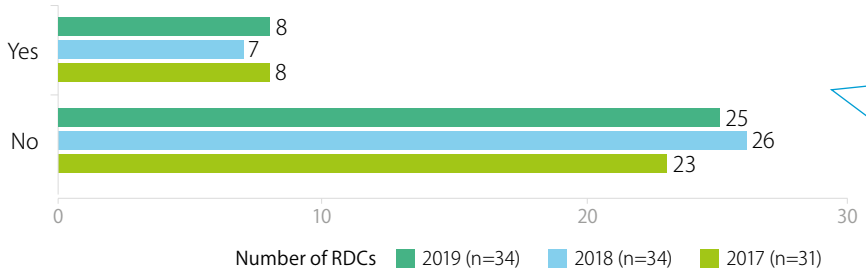


Half of the RDCs make datasets available without closure periods

Half of the surveyed RDCs do not have closure periods, i.e., the data are made available immediately after they are received and prepared (see Fig. 14). Sixteen RDCs reported imposing closure periods on at least parts of their data in 2019. Five RDCs report fixed closure periods ranging from six months to no more than two years. Closure periods at all other RDCs are not fixed but depend on certain requirements, for example, a research project's end date, or other specifications by funding organisation or data givers. Overall, there has been a certain dynamic in the way closure periods have been implemented; the absolute number of RDCs that use closure periods for datasets saw a slight increase.

Fig. 15: Fees for data access at RDCs

Do you charge fees for data access for research purposes?



A large majority of RDCs does not charge fees for data provision

In addition to closure periods, user fees can obstruct data use. A majority of RDCs do not charge any fees at all: out of 34 RDCs, 25 reported that they had not charged any fees in 2019 (see Fig. 15). Compared to the previous year, the differences were very slight. The fees reported by eight RDCs were mostly in the two-digit or lower three-digit euros range. Many RDCs offer discounts for PhD theses.[8] RDCs vary in the way they charge fees: some charge per dataset, or data access path, while some charge for data use per survey year. Some charge fees to account for the effort that goes into certain data preparation measures, such as specific anonymisation measures, or additional support services. It is safe to assume that these very low fees contribute little to covering the RDCs' expenses. However, in the sense of a token fee, they help prevent "fake requests" and ensure that the data are used only for their designated purposes.

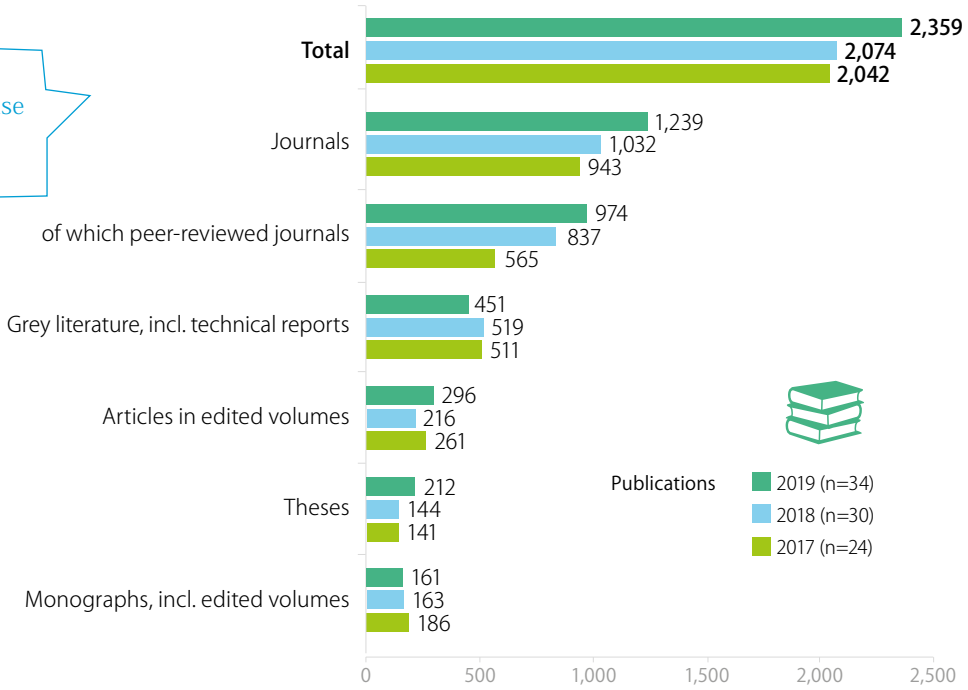
8 The Federal Statistical Office has been waiving the already reduced costs for data use by young researchers since June 2020. In the entire official statistics community, data access for bachelor's, master's and PhD theses is now free-of-charge.

Research output based on RDC research data

Research output is a key indicator for successful data use. The number of publications, in turn, is an indicator of this parameter. Measurements must take into account qualitative differences between forms of publication. Articles in peer-reviewed journals, for example, must be weighed more than articles in journals that are not. Articles in academic journals were the publication form reported most by data users in 2019. The number of peer-reviewed journals, here, is particularly high, and also saw the highest growth rates.

Fig. 16: Publications based on research data provided by the RDCs

There is a continuous increase in the number of publications



In 2019, 34 RDCs reported a total of 2,359 publications (see Fig. 16), which were based on their available research data. It must be noted that many researchers neglect to notify the RDCs about a publication or send a copy. Moreover, despite widespread use of persistent identifiers for research data by RDCs, many researchers do not yet make use of data citations in their publications. Since many RDCs do not have the resources to continuously gather data on citation, it is safe to assume substantial underreporting here. The fact that researchers may use datasets from several RDCs also contributes to this fact.

Even though calculating the number of publications that are written based on the data made available by RDCs is not without problems, there has been an increase in publications based on research data provided by the RDCs over the years. As mentioned above, this is particularly true for journal articles. Furthermore, the number of theses working with RDC data also grew significantly over the years.[9] The number of monographs, including edited volumes, have remained fairly constant. Interestingly, the percentage of grey literature and technical reports has been declining compared to previous years.

9 Considerable underreporting is also to be expected for theses, since particularly bachelor's and master's theses are hard to find and usually only recorded when data users make an effort to do so. In addition, authors of PhD theses tend to work within larger projects and not to register data use directly with the RDC. The number of theses written using RDC data is likely much higher.

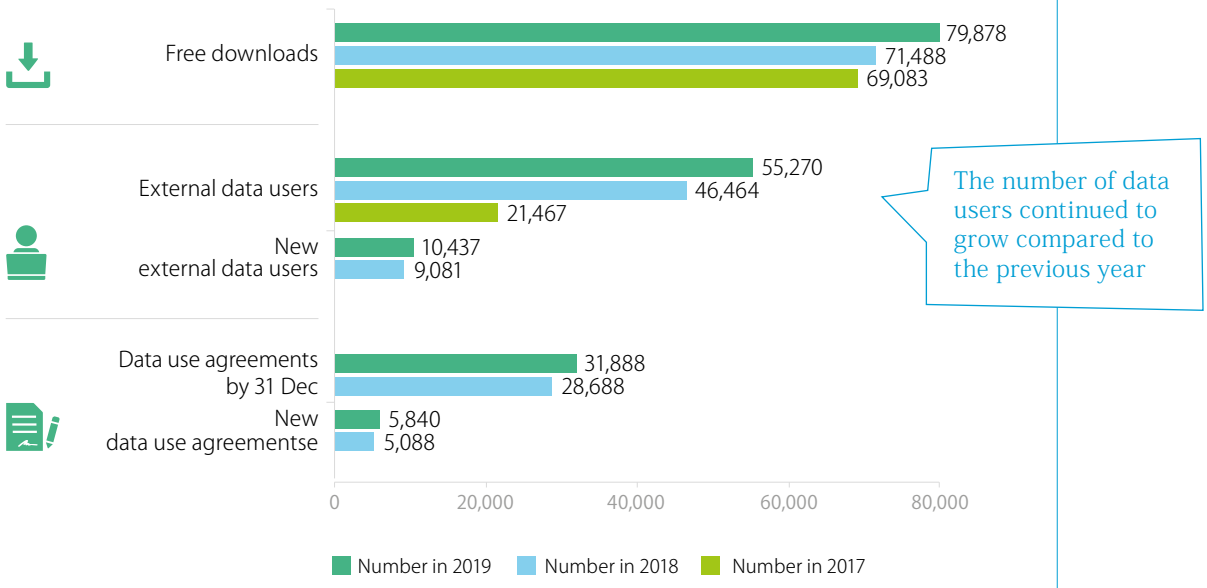
Scope of data use

The scope of data use is another key indicator for the relevance of RDCs in the research landscape. One central variable is the number of datasets that were retrieved from the RDCs. Additionally, the number of researchers that benefitted from RDC services is also instructive.

Since some RDCs are highly integrated into their parent institutions, while others have a strong service infrastructure and external orientation, putting the usage figures into context is not trivial. One of the core tasks of the RDCs is to provide researchers with comprehensive and flexible data access, which they aim to continuously expand and improve. Due to specific data protection regulations and other legal provisions, researchers are offered several different data access paths. Owing to this flexibility in access paths and the differences in how data on contracts, projects, and data users are gathered by the RDCs, it is difficult to determine a precise number of data users. In previous years, it was not possible to rule out the double counting of contracts, projects, or between access paths.

Getting a clear idea of this diversity has proven to be complex. For this reason, the respective survey questions were redesigned and split into four partial indicators: number of downloads, external data users, data use contracts, and the number of surveyed users and datasets (see Fig. 17).

Fig. 17: Indicators for the number of data users



Downloading data

79,878
Free downloads
n=17

Overall, 17 RDCs make datasets available as free downloads. At the majority of RDCs, downloading requires prior registration. Some datasets are freely downloadable without prior registration -- making it more difficult to identify users. Therefore, not all forms of data access and users can be quantified. Furthermore, not all data access paths have the technical means needed for gleaning usage statistics. All this contributes to an undercounting of data users. The 11 RDCs which can provide information on user numbers reported 79,878 downloads of open datasets in 2019.

The standard case, however, is that RDC-held research data are made available only after users have registered or signed a contract.

External data users

55,270
External data users
n=29

10,437
New
external data users
n=27

In the 2019 reporting year, the RDCs had a user base of 55,270 external data users. External users are data users that are not affiliated with an RDC, or an RDC's parent institution. 29 RDCs were able to provide data on the number of external data users.

The number of external data users increased by 10,437 persons in 2019. Twenty-seven RDCs were able to provide data on additional data users in 2019. Overall, this means that the number of new data users in 2019 might be higher.

Data use agreements

31,888
Data use agreements
n=31

5,840
New
data use agreements
n=33

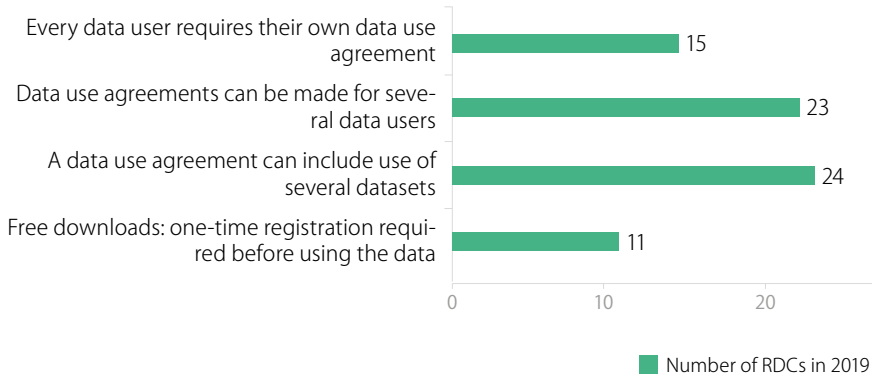
For data protection reasons, contracts on data access and usage contain explicit references to research projects (purpose limitation), i.e., a separate contract must be drawn up for every research project using the data. However, there are no formal templates for such contracts. How RDCs design their contracts is governed by the freedom of contract, and the contractual depth is determined by legal provisions and requirements. Access to official statistics data, for example, is legally restricted by a string of laws and regulations. Access to survey data is also subject to data protection regulations. This applies particularly to sensitive personal data. Other data, on the other hand, including regionalisation and land use data, are openly accessible for some purposes and subject to licensing for others. This diversity is also reflected by the contract design used at RDCs. This is true for the data themselves as well as the signatory parties as contracts can be with individuals, projects, or entire research facilities. They partly cover entire data troves, collections of studies, or individual datasets.

On the cut-off date for this publication (31 Dec 2019), 31 RDCs reported 31,888 existing data use agreements in the RatSWD's research data infrastructure. Thirty-three RDCs signed 5,840 new data use agreements in 2019.

Fifteen RDCs reported that each dataset required an individual data use agreement. Twenty-three RDCs make contracts covering several users, eleven of which restrict use to specifically stated individuals. On average, contracts provided access to two persons. Seventeen RDCs sign contracts at a project level; nine RDCs enable data access for entire institutes (see Fig. 18).

Fig. 18: Contract design

Which of the following applies to your RDC? *(Multiple answers possible)*



Contract design varies strongly: contracts can cover one or several persons or datasets

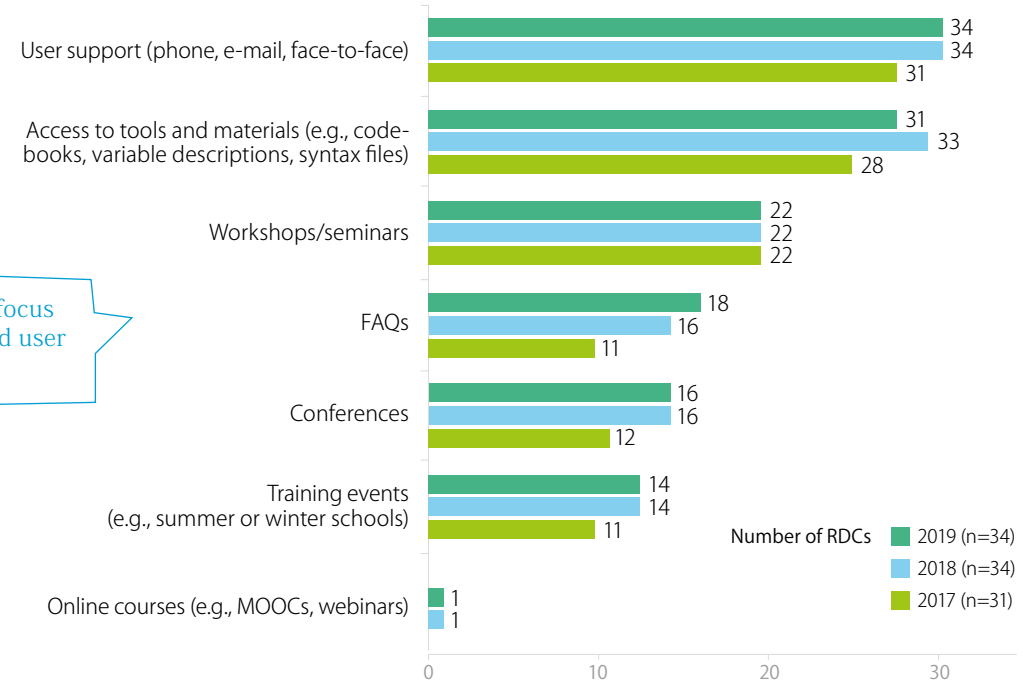
At 24 RDCs, data use agreements govern access to several datasets. In 13 cases, only one designated person may use them; 14 RDCs extend access to include several people. On average, contracts covered slightly more than two persons. Project-level contracts were used by eight RDCs. With only one RDC, institution-level contracts are an exception.

5 Services for data users

This chapter is dedicated to the user-focused pillar of the RDCs' activities. As shown in Chapter 2, the structure of the RDCs allows for a considerable amount of active research on behalf of its academic staff (see Fig. 5: Independent in-house research by academic staff). The high share of staff involved in active research at the RDCs is the foundation for their user orientation, because the staff are uniquely informed about the potential and the pitfalls of the available data and the service demands of users. User services are therefore an important part of the work of research data and data service centres. Direct user support regarding data access and datasets is equally important for the preparation of datasets and creation of tools and materials. The following chapter will look at RDC user services in greater detail.

Fig. 19: User services at RDCs

Which user services does your RDC provide?
(Multiple answers possible)



All RDCs focus on tailored user support

All RDCs offer customer support to external researchers, which is provided by designated RDC staff via telephone, email, or face to face (see Fig. 19). RDC staff are entrusted with the important and complex task of providing researchers with tailored support. This service is made possible by the fact that RDC staff have ample experience in working with their in-house datasets and are thus uniquely familiar with the data's analytic potential.

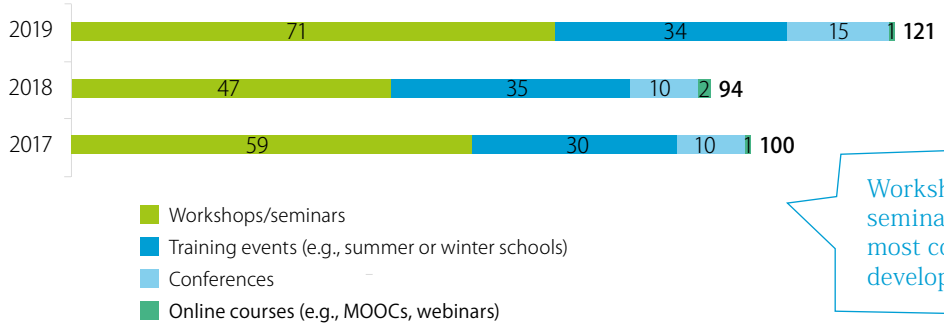
Compared to the previous year, the survey did not register any changes regarding the number of workshops/seminars, conferences, or training events. The number of available tools and materials, such as codebooks, variable descriptions, and syntax files, decreased slightly; the number of FAQs, on the other hand, saw a slight increase. One RDC offers online courses to its users. Additionally,

a small number of RDCs provide online support, such as online tutorials or message boards. The aim of all these services is to enable data users to better exploit the analytic potential of each dataset (see Chapter 6).

Skills development

In addition to the services for data users, RDCs offer an array of skills development measures in various formats aimed at making researchers more qualified to work with the datasets. A total of 24 RDCs offer 121 skills development measures, increasing their contribution to user qualification by 27 training events from the previous year. Workshops are a particularly effective ¬¬ way to train users in relatively small groups: RDCs hosted 71 workshops and seminars this year – an increase of 24 events compared to the previous one. While the number of training events and online courses remained the same, there has been a slight increase in the number of conferences.

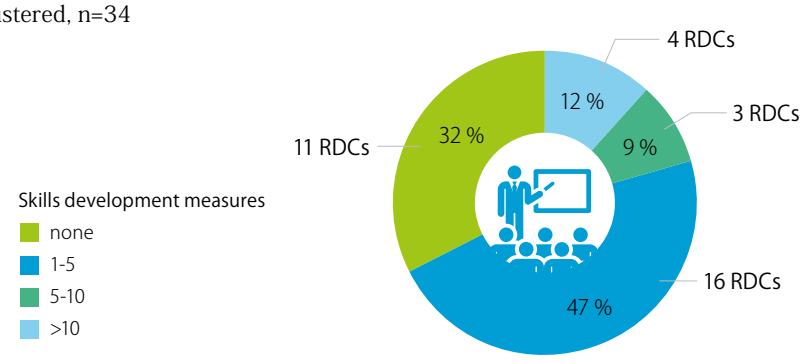
Fig. 20: Skills development measures
How many of the following skills development measures does your RDC offer?
(Multiple answers possible)



Workshops and seminars are the most common skills development measure

Skills development measures are not evenly distributed across RDCs (Fig. 21). Eleven RDCs do not offer any at all. The majority of accredited RDCs, however, offer between one and five skills development measures. Remarkably, four RDCs hosted more than ten skills development events.

Fig. 21: Number of skills development measures at RDCs
Clustered, n=34



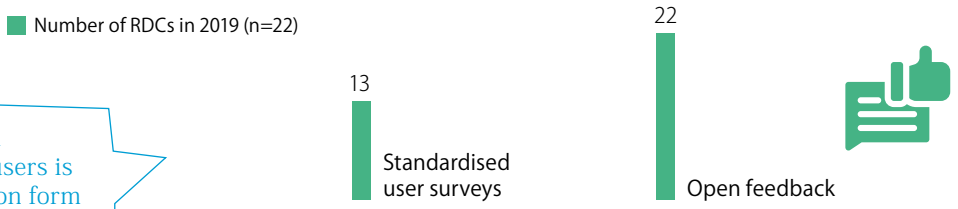
The number of skills development measures per RDC varies strongly

Service quality assurance

Most RDCs use process-integrated measures to assure the quality of their services. A total of 13 RDCs use standardised user surveys, and 22 RDCs use open user surveys to collect feedback for quality assurance. Some RDCs also collect feedback via their website, or via questionnaires following training events. The frequency with which standardised user surveys are conducted varies significantly. Seven RDCs report conducting these surveys continuously, while two RDCs do it annually, and four RDCs less than once a year. The results are collected, discussed, evaluated, and, if possible, implemented. Some RDCs publish results on their websites and use this feedback to guide the development of the infrastructure, or for internal reporting. Open feedback mechanisms are diverse, ranging from bilateral talks with users to quality assurance mechanisms of the RDC's scientific advisory board or parent institution.

Fig. 22: Quality assurance through user surveys

How are services evaluated, and their quality secured?
(Multiple answers possible)



Evaluating open feedback from users is the most common form of quality assurance

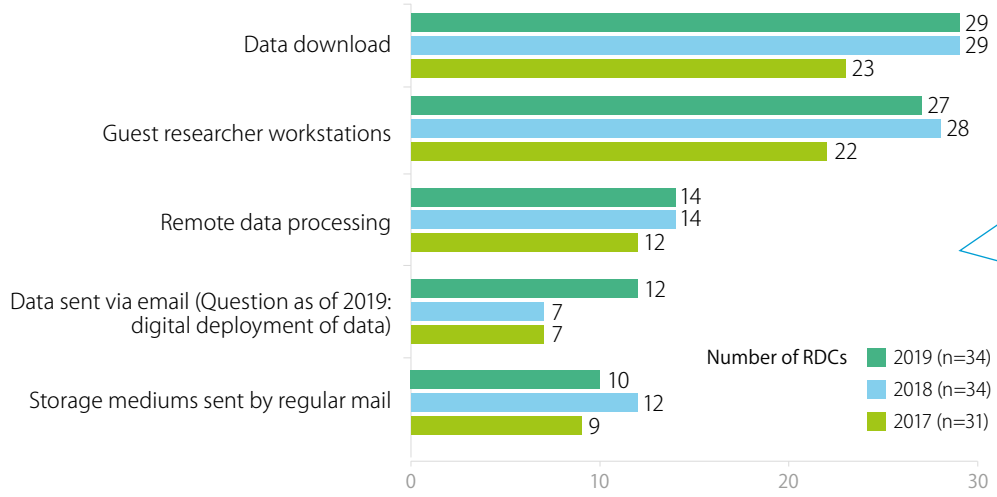
To summarise, it is found that the extent of quality assurance is as varied as its design. The common ground is that, independently of the type of quality assurance, the RDCs value feedback and aim to implement suggestions about their events, or to optimise their services.

Data access paths and data formats

The RDCs provide access to data through two main access paths: (1) on-site use, and (2) off-site use. On-site use (1) involves researchers working on guest researcher workstations (*Gastwissenschaftsarbeitsplätze*, GWAP), enabling them to access confidential and sensitive data (see Fig. 2 for a map of guest researcher workstations in Germany, and info box 2 p. 36 for a definition of guest researcher workstations, or GWAPs). RDCs also enable their users to access data off-site (2). While it depends on the individual dataset, off-site use involves transmitting a dataset to researchers via post or digital means, who can then work with them in their own infrastructure. This often requires signing a data use agreement. It also includes remote data processing, such as the use of a VPN connection or access through a designated online portal.

Fig. 23: Data access paths offered by RDCs

Which data access paths does your RDC offer?
(Multiple answers possible)

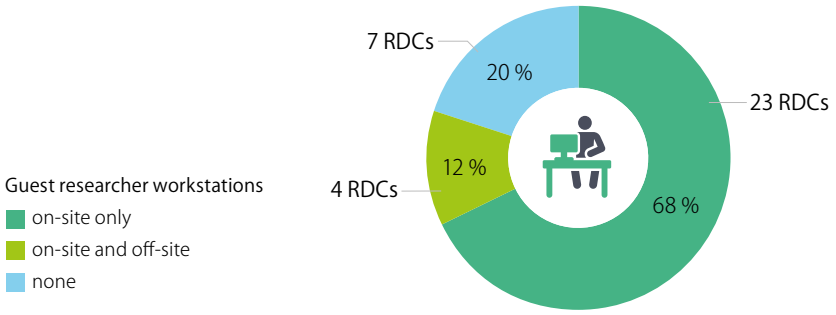


Deploying data digitally is becoming more important

More than three quarters of RDCs (27) offer guest researcher workstations. Four of these enable researchers to work with their data on off-site guest researcher workstations, outside of their own institution (see Fig. 2, p. 14 for a map of guest researcher workstations in Germany). Compared to the previous year, however, one RDC has stopped giving access to sensitive data (see Fig. 23).

Fig. 24: Guest researcher workstations in 2019

Which data access paths does your RDC offer? (n=34)



Guest researcher workstations enable access to sensitive data

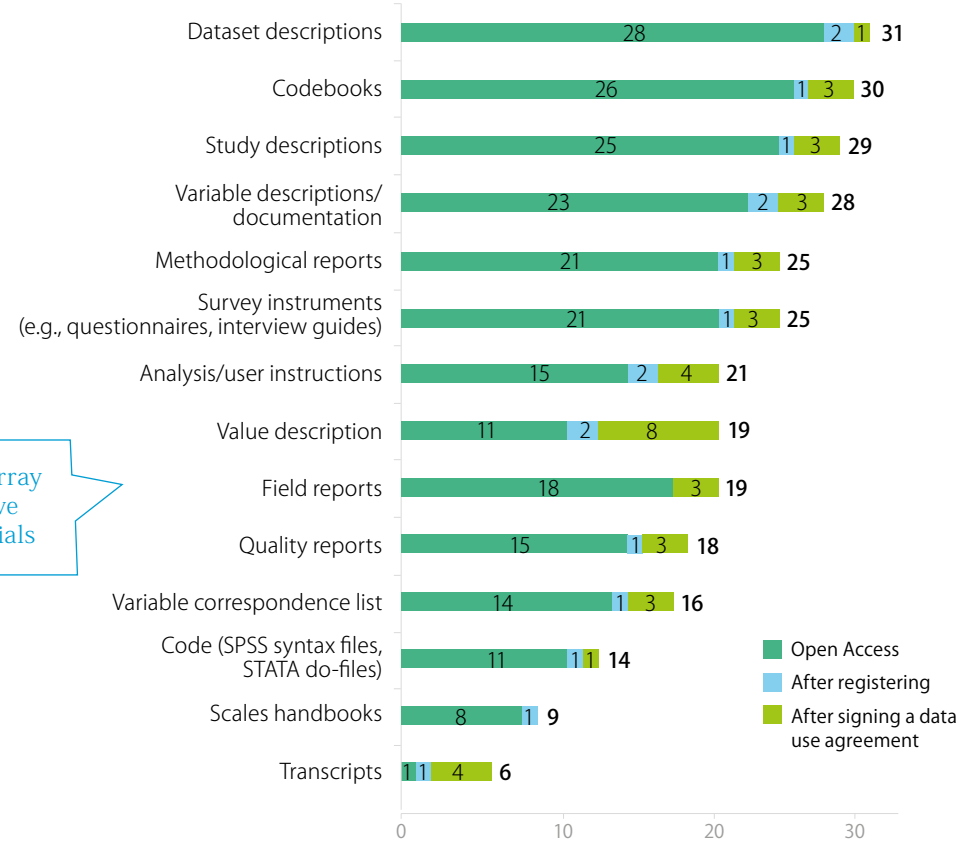
For off-site data use, RDCs offer a variety of access paths. A total of 29 RDCs enabled downloads of datasets for off-site use as of last year. Data access via (controlled) remote data processing is still offered by 14 RDCs, e.g., via a remote connection or by submitting analysis scripts. Data transmission via email is available at 12 RDCs and via data storage mediums sent by regular mail at ten RDCs. RDCs typically opt for either analogue or digital deployment—sending data either via mail or email. Only one RDC offers both. The numbers show that data deployment via email is gaining ground compared to the previous year (7), and that postal shipping is declining.

Provision of tools and materials

Data users benefit from dataset-specific tools and materials to gain a better insight into the available datasets and thus to improve their ability to realise their research project using those datasets. Comprehensive dataset descriptions or codebooks, for example, can help decide whether a dataset is suitable for a research project ahead of applying for it, which also reduces the number of applications for datasets that must be reviewed. This benefits the users as well as the RDCs—the latter can reduce processing of unnecessary contracts. The following figure gives an overview of tools and materials offered by RDCs and how they are deployed.

Fig. 25: Provision of tools and materials in 2019

Which tools and materials do you offer for which data access path? (n=33)
(Multiple answers possible)



RDCs offer an array of comprehensive tools and materials

Almost all RDCs provide users with information and documentation for each of their datasets. They vary according to the RDCs and the datasets. Tools and materials for working with the data tend to be open access,[10] enabling researchers to gather information before accessing the dataset. The most common tools include dataset descriptions, codebooks, variable descriptions, variable documentation, study descriptions, and methodological reports.

10 Open access means that scientific literature and other material are available openly online.

RDCs often make available dataset-specific tools, including code (e.g., do-files, syntax files), technical metadata descriptions (e.g., XML files), web applications (e.g., metadata search systems), and theory-focused documents, which inform users about the contextual background of certain questions and variables. Access to tools is often restricted until after a data use agreement has been signed, especially code and transcripts. The reason why access to some tools is restricted until after registration or signing a contract is due to the type of tools: e.g., some tools contain sensitive information that are restricted for data protection reasons, or a syntax that is only useful when the appropriate dataset is on hand. In some cases, data are provided only after a formal review to guarantee non-commercial use.

Tools and materials are generally available in an open access format to download via the RDCs' websites and without prior registration. If the research data are provided to users in a package, this package tends to contain all the available tools.

Moreover, some RDCs offer other support tools, for example, to help with citing research data, or by providing information on board notes, seating plans, or interview protocols in qualitative research. RDCs also use new forms of communication to support data users in their work, including video tutorials on working with data platforms, or interactive metadata portals.

Time period between the application and the data access

Data eligible for online access are usually available immediately or within a few hours (e.g., after a simple registration process, or downloaded directly from the data catalogue by the users). Sensitive, less anonymised data may not be downloaded online.

The time period between signing the data use agreement and the transmission of the data by the RDC can range between one hour and several weeks. Users may encounter longer processing times when requesting specially prepared data, when the data require it, or where the data release requires special permits, e.g., a separate review for data use abroad. More than half of RDCs make data available within a week after a contract has been signed (19 RDCs), and 12 RDCs within two to four weeks. Only one RDC takes longer than four weeks.

Info box 2:

Data access paths and data access formats

Guest researcher workstations

These are specially secured workstations at research data centres (RDCs) enabling researchers to access data that are at least formally anonymised. It is common for guest researcher workstations not to have uncontrolled internet access and to disable local saving of files.

Remote data processing

This data access path enables researchers to perform data analysis at RDCs without being on location. Researchers submit analysis scripts to a RDC's staff, which they write at their respective workplaces (sometimes using a structured dataset to test their code). Depending on data sensibility and/or legal regulation, the resulting files are checked at the RDC (output control) before being sent back to the researchers. This method is called remote data processing or remote access.

Scientific use files (SUFs)

SUFs are research datasets that are de-facto anonymous datasets but still have considerable analytic potential.

Campus Files (CFs)

Highly anonymised research datasets created for academic teaching purposes.

Public Use Files (PUFs)

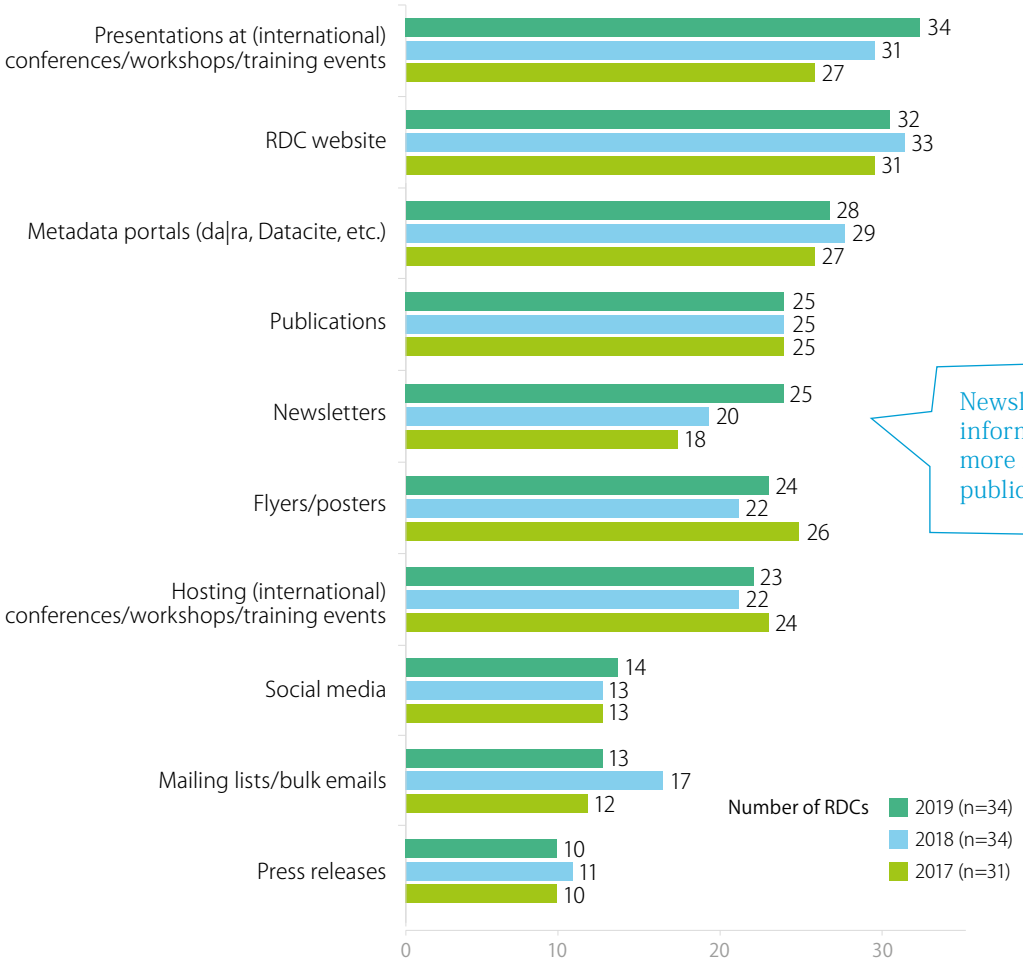
Anonymised research datasets without use restrictions that can be shared for non-academic purposes.

Publicising the research data

The RDCs have an interest in the extensive use of their data. They use a multitude of ways and channels to publicise their data offering, highlight its potential for research, and stimulate data use. These are spelled out in greater detail below.

Fig. 26: Communication channels used to publicise RDC data

How and through which channels do you publicise RDC data in the scientific community? *(Multiple answers possible)*



Newsletters and face-to-face information are becoming more important for publicising RDC data

Self-promotion is the most important channel, particularly through RDC websites, newsletters, as well as presentations at (international) conferences. Metadata portals like da|ra and DataCite also play a central role in publicising research data and attracting more users. Widespread use of these services is a promising basis for improving the accessibility (and findability) of RDC data for other disciplines at the national and international level. This is followed by putting out publications and hosting training events and workshops. Publishing press releases and content on social media remains less widespread. It will be interesting to track how this develops in the following reporting year, seeing as face-to-face events have been severely limited due to the COVID-19 pandemic, and whether online channels will gain in importance.

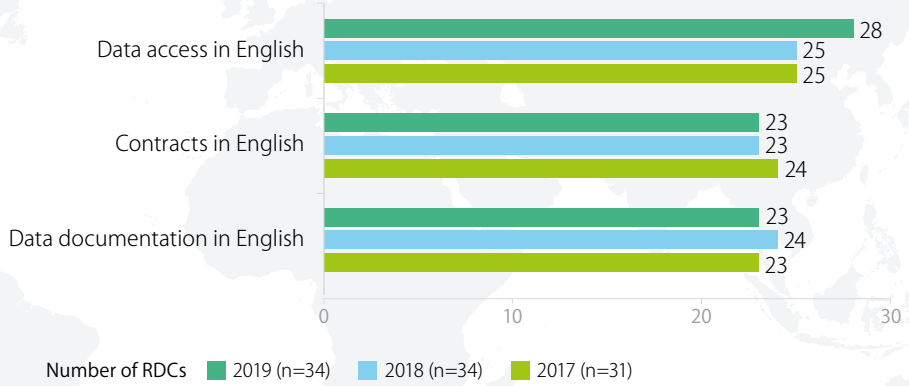
6 Current development of the research data infrastructure in the social, behavioural, and economic sciences

Internationalisation

The increasing global interdependence of economies, societies, and policy is giving rise to new fields of research, which can only be addressed within the framework of international scientific cooperation. On the one hand, a prerequisite for this is that international researchers can access national-level data. On the other hand, there is also an increasing demand for international microdata, which improves comparative analysis.

This internationalisation of the research landscape is also an everyday practice at RDCs. Responding to the demands from within the scientific community, research facilities and research infrastructures have an increasingly international focus. RDCs have created access paths and data documentation in English to cater to the international research community in Germany.

Fig. 27: Support for international researchers
How do you support international researchers?
(Multiple answers possible)

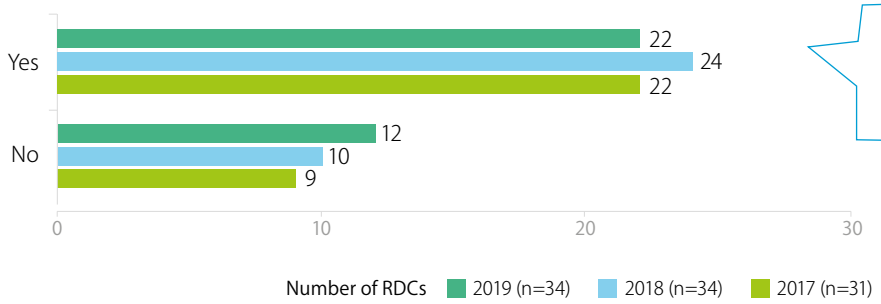


Three additional RDCs now offer access in English

Overall, 28 RDCs now facilitate international data access by providing access paths in English. While the number of RDCs remained the same, three additional RDCs now do this. This indicates a real increase in international initiatives, not simply a result of the accreditation of new RDCs. At 23 RDCs, this includes data documentation; 23 also offer user agreements in English (see Fig. 27). Furthermore, international users are provided with tailored user support and communication via email and phone in English. Additionally, RDCs provide translations of contracts, newsletters, user workshops, conferences, and training events in English. Almost every RDC (31) offers English-language user support; beyond this, 27 offer not one but several support services in English.

Fig. 28: International contacts of RDCs

Does your RDC have contacts to international research facilities?

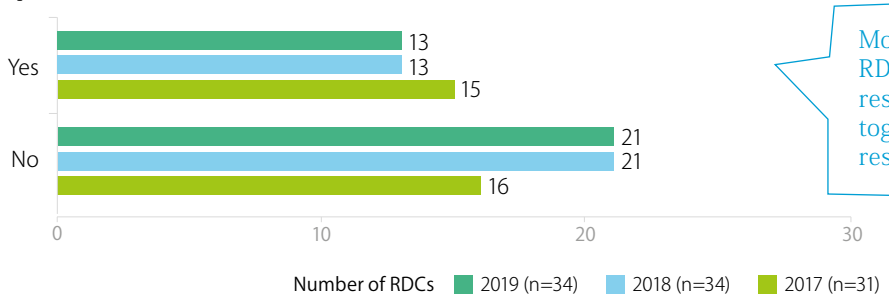


Two out of three RDCs continue to maintain international contacts

The share of international data users varies greatly across RDCs. In total, 22 RDCs have contacts to international research facilities (see Fig. 28, see Chapter 'Cooperation and Research Activities' on p. 16 for more information on cooperation at the national level). This includes collaborations between individuals, international research institutions and consortia, as well as international cooperation and harmonisation of procedures, for example, addressing issues of data access and privacy at the European level.

Fig. 29: RDCs with close international research partnerships

Does your RDC maintain international research partnerships? (i.e., international researchers working together on specific issues)



More than a third of RDCs continue to have researchers working together in international research partnerships

In addition to supporting the global scientific community in using RDC data, providing English-language information and materials, and maintaining international contacts, the RDCs work together with international researchers on specific issues in international research partnerships. In the 2019 reporting year, 13 RDCs reported that they maintained international research partnerships to work on shared subjects (see Fig. 29), which are also very diverse and include joint projects, working groups, and other forms of collaboration with European and international facilities such as universities, RDCs, data archives, and research institutes.

Innovation and improvement of the research data infrastructure

2019 saw a broad range of innovations in and improvements of the research data infrastructure. In general, most RDCs reported that they continuously expanded their range of data, the availability of new datasets, and data access paths.

A striking innovation was the introduction of new data access points, such as guest researcher workstations (GWAPs), at a national and international level and with a high safety standard—often combined with remote access solutions.[11] Many RDCs continued to heavily invest into expanding the infrastructure. Other areas where improvements were made included data linkage, documentation in the form of metadata, RDC websites, and application procedures.

Further development of the research data infrastructure

The following segment cites a range of issues in which RDCs indicated a need, or an interest, in deepening knowledge exchange. Data protection continued to be an important issue throughout the 2019 reporting year, with many RDCs indicating an interest in knowledge transfer and support. Other key points included exchanging information on user agreements and data transfer agreements, authoritative legal counsel on data protection and copyright issues, as well as secondary research data use and anonymisation measures against the backdrop of the GDPR.

In the field of *research data management*, five key issues were identified as particularly relevant for the future of RDCs:

1. The RDCs would like to exchange information on methods and tools to deal with user requests, including application and contract management, and automation of administrative processes and archiving. They also mentioned developing criteria for data quality, measures to increase secondary use, portfolios and tools to support data preparation during the research process, skills development of young researchers, user training, and request management systems.
2. The RDCs also expressed an interest in discussing methods and tools for data harmonisation, and ways to facilitate information flows, such as structured metadata – from data collection until their end use.
3. On the issue of data access, the RDCs plan to exchange knowledge and ideas on alternative and innovative data access paths, expanding the 'RDC-in-RDC'-approach, remote data processing, (automated) output control, and the implementation of remote access solutions. They also called for comparing the pros and cons of being classified as a scientific institution.
4. On the issues of sensitive and qualitative data, the RDCs mentioned the following points: data protection measures as well as the collection and provision of qualitative data, remote access to sensitive data, and the combination of research data and various mixed-method approaches.
5. Expanding the research data infrastructure and increasing networking among the RDCs continue to play an important role. Responses also raised the issue of how to deal with new tasks, for example, RDCs in their roles as data trustees, and guest researcher workstations for external data.

The RDCs also wish to foster knowledge exchange regarding the following issues: harmonising processes among the RDCs, deployment of tools for dataset search and for communicating with users (e.g., user online portals), standards and automation technology for controlling data exports in remote data processing and guest researcher workstations, certification procedures, and interconnected guest researcher workstation.

11 GWAPs only provide the IT equipment for accessing microdata. Access to the internet and other sources is not possible. Users go through physical access control and are monitored to ensure that the contractual restrictions are adhered to; trained staff in charge of keeping the safety standards make sure that only authorised persons access the facilities. This includes physical safety measures such as window covering and barriers within rooms marking off individual workstations, i.e., computer screens. The design of these facilities varies according to each data provider's security requirements.

7 Special topic: data linkage

The special chapter 'Data Linkage" aims to give an overview of the extent to which linking datasets is possible and commonly used at RatSWD-accredited RDCs. Data linkage here means so-called record linkage, i.e., directly joining data together from at least two datasets via a distinct identification number (ID), or a distinct combination of attributes, respectively. This can be used to bring together information on individuals, businesses, or region codes and geospatial data, which were collected in different surveys and/or official statistics.[12]

Significance and potential of data linkage for the RDCs

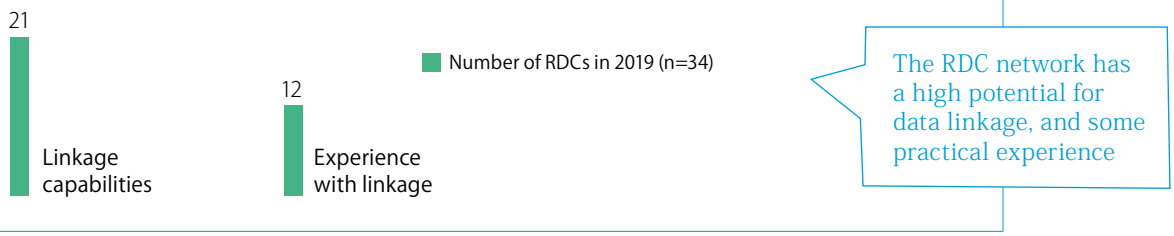
Data linkage is an important topic at many RDCs, because, by joining datasets, the analysis potential of each one is enriched. By doing so, data linkage enables RDCs to potentially address certain research questions that they would have otherwise been unable to because of missing information.

A majority of RDCs (21) identified possibilities for linkage within their data troves. At 18, more than half of research data centres have at least[13] 71 datasets in their portfolios that are suitable for data linkage, because they contain at least one unique identifier or feature set. The individual data in those datasets contain identifiers such as social security number, personal ID number, company registration number, employer number, tax ID number, or geospatial data.

However, only 12 RDCs reported having concrete experience working with data linkage. Most of them usually linked several datasets; overall, a total of 29 datasets were linked in this way. Almost two thirds of these linkages (19 datasets) involved other RDCs.

One reason why linkage is not used more often lies in the restrictions that obstruct or prevent such projects.

Fig. 30: Capabilities and experience in data linkage



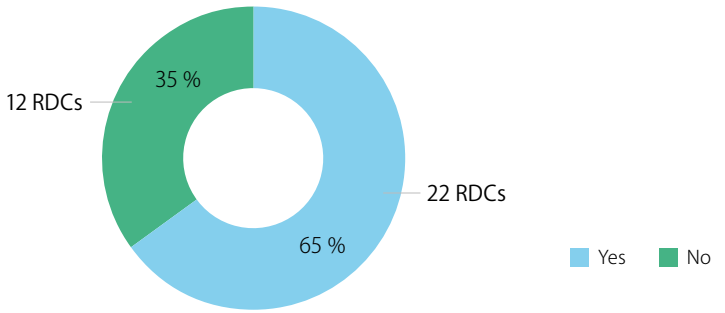
12 This chapter is not concerned with 'statistical matching', which uses data from several datasets and integrates them by using propensity scores.

13 This number is not conclusive, however, because some RDCs list every available dataset, while others make more general statements on datasets (e.g., 'all datasets', or, more generally, 'economic statistics' and 'environmental statistics').

Fig. 31: Relevance of restrictions

Are legal or other restrictions relevant for your RDC's capacity for linking datasets?

Almost two thirds of RDCs report restrictions on data linkage

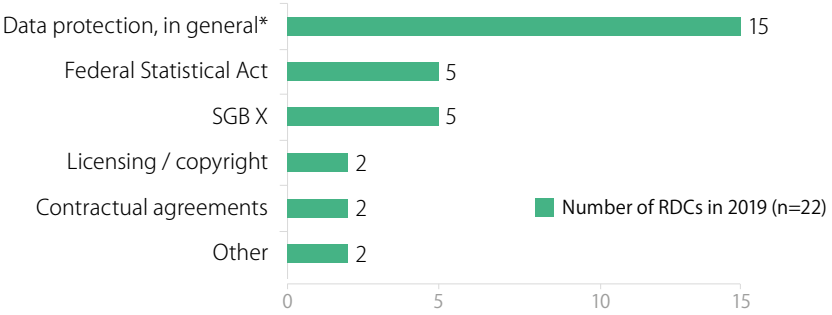


Twenty-two RDCs, which is almost two thirds, report that data linkage is subject to restrictions, which are often legal in nature.[14] Fifteen RDCs point to data protection regulations at the European level (GDPR) as well as at the federal (BDSG) and state level. Any personal data is subject to these regulations, because their processing and linking is governed by law, e.g., requires the informed consent of study participants, or other legal requirements. The legal restrictions of the Federal Statistics Act and the Social Insurance Code (SGB X) are each specifically mentioned five times. Only two RDCs mention restrictions that stem from contractual, licensing, or copyright issues.

Fig. 32: Relevant external restrictions

Which legal or other restrictions are relevant when linking datasets at your RDC?
(Evaluation of an open question, multiple answers possible)

Record linkage at the RDCs is restricted mainly through data protection regulation



* Respondents either generally spoke of 'data protection', or specifically named GDPR, federal, and state-level data protection lawse

This means that, while linking data may generally be possible, the RDC in question may not permit it. It makes evident that data linkage is heavily restricted, occasionally enabling some projects but presumably preventing many conceivable others.

14 These data are based on open questions, i.e., the responses were categorised later.

Levels of data linkage

Linking data is possible at the individual, company, or regional level. This next section will focus on the linking of individual, firm, and company-level data, because they raise similar challenges regarding data protection. We will then look at linking regional data, which requires different procedures, but is quite safe regarding privacy concerns.[15]

Linking individual data such as personal or company data

When research data from the social, behavioural, and economic sciences are collected and subsequently shared for secondary use, they do not typically contain names, addresses, or other unique identifiers, allowing for reidentification of research subjects. This is either guaranteed to respondents during surveys and interviews, or, and such is the case with process-produced official statistics data, safeguarded through standard anonymisation procedures. A further possible characteristic of empirical datasets is that they are pseudonymised. The latter data can also be linked.

However, linking datasets requires using the same identifiers. Planning for such research projects therefore usually involves settling for a common identifier. However, not all RDCs have all identifiers at their disposal: using the social insurance number, for example, is not allowed if the participating institution is not a social insurance-related institution, or if a legal exception is not in place (§ 18f. SGB IV).

Identifiers which can serve to join datasets together on these levels are available as data on individuals, companies, firms, or households in RDC datasets.

Ten RDCs have at least 30 datasets, which are potentially linkable at this level[16], Four of these RDCs are capable of linking data at the company level and also have the company registration number to do so. At seven RDCs, bringing together datasets on individuals may be possible, for example, by using the – partially pseudonymised – social insurance number, which five RDCs have available for at least 19 datasets.

Identifiers at the household level (or the level of the Bedarfsgemeinschaft, a household entity specific to the German social code) are rarely suitable for linkage (two mentions).

Only six RDCs actually linked data by means of attributes belonging to individuals, either using the social insurance number (three RDCs for 15 datasets), a Merge ID (two RDCs for three datasets), or a feature set, as done by only one RDC (name and address data, linking of two datasets). At the firm and company level, two RDCs linked together four datasets, two of which used the firm or company registration number.

This makes the individual the most important identifier, followed by firms and companies.

15 Data protection issues should only arise here if small numbers of certain groups of people are available for individual regions, facilitating reidentification of individuals.

16 This number is not conclusive, however, because some RDCs list every available dataset, while others make more general statements on datasets (e.g., 'all datasets', or, more generally, 'economic statistics' and 'environmental statistics').

Linking data via geospatial data and other regional identifiers

Nine RDCs reported 31 projects that could involve linking data at the regional level. Two thirds of these RDCs (6) reported that linking their datasets could be possible using spatial data. Georeferenced data are the most mentioned and are available as unique IDs[17] for at least 29 datasets.[18] This shows that some RDCs could conduct a high number of linkage projects using region-level data. In many cases, georeferenced data are reported as a possible level for linking datasets, while some mention other regional indicators (e.g., addresses, parishes, counties, regional planning units, state).

In fact, RDCs rarely performed linkages using spatial identifiers to date. Only one RDC linked a total of three datasets using georeferenced data.

Professions and schools

Some RDCs mentioned the possibility of linking datasets at the level of professions and schools. However, out of all the datasets suitable for linkage, this type has never been used and has generally played a subordinate role.

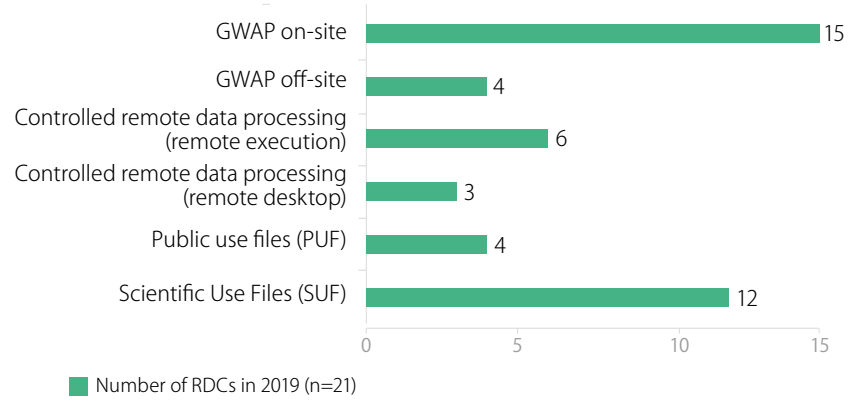
Access paths for data linkage

The RDCs offer different access paths to linking datasets.

Fig. 33: Existing access paths

Which access paths does your RDC offer for linking datasets?
(Multiple answers possible)

RDCs most often offer on-site guest researcher workstations for linking datasets



17 Other IDs that were reported were county and parish indicators (four times) and regional indicators (twice). Other indicators that were mentioned once include Inspire Grid, address, regional indicator, regions used in spatial planning.

18 This number is not conclusive, however, because some RDCs list every available dataset, while others make more general statements on datasets (e.g., 'all datasets', or, more generally, 'economic statistics' and 'environmental statistics').

Out of 21 RDCs stating that they have datasets with at least one unique identifier, 15 RDCs make these linkable datasets available via on-site guest researcher workstations (see the info box on p. 36 for more detailed information on access paths and data formats). Four RDCs provide an access path to data linkage via off-site guest researcher workstations (see Chapter 'Data access paths and data formats', p. 32). Access via GWAPs is therefore the most common access path. Twelve RDCs offer datasets in the form of scientific use files, which researchers can analyse off-site from their own desks. Only four out of 21 RDCs make data available as public use files, which can be used, for example, in university teaching.

Using controlled remote data processing, six RDCs make data available for linking via either remote execution or remote desktop procedures, and a total of seven RDCs offers remote data processing. Making data available via guest researcher workstations and controlled remote data processing accounts for the specific requirements of this type of data use: the data contain unique identifiers that can be traced back to statistical units in numerous ways, making them particularly sensitive.

To summarise, the linked datasets often provide a rich source of information, but they are typically only accessible via guest researcher workstations. Generally, though, linking datasets (currently) still plays a subordinate role at RDCs. The special requirements of data linkage regarding the privacy of respondents and adherence to other legal requirements tend to be an obstacle to a more widespread use of data linkage. However, based on this initial survey, it is not yet possible to say whether the access paths to linked datasets will become more diverse if, for example, there is an increase in demand.

8 Complaints management

One of the key tasks of the RatSWD is to assure and improve the quality of RDC services. Since its inception, the RatSWD has acted as a dedicated point-of-contact for complaints relating to RDC data and services. In addition to overseeing the annual monitoring process, of which the present activities report is one outcome, the RatSWD's monitoring commission also handles complaints put forward by research data users.

The RatSWD set up a complaints office at the RatSWD business office to professionalise complaints management and make it more transparent. The complaints office ensures a swift and professional response to complaints and feeds the results back into RDC processes to further improve the data infrastructure.

If data users become aware of major shortcomings in the data services of an accredited RDC, it is recommended they first approach the RDC directly to try to find a solution. If the problem cannot be resolved, users may direct their concerns to the complaints office. The complaints office's mandate is limited to issues concerning compliance with the RatSWD accreditation criteria. The RatSWD is not responsible for delays during everyday procedures or for staff conduct at RDCs. Complaints of this nature should be directed to the RDC in question.

For more detailed information about the procedures, see the RatSWD Output 8 (5)[19] or the updated version of that output on the German Data Forum (RatSWD) website[20].

Current complaints procedures in the 2019 reporting year

The RatSWD did not receive complaints during the 2019 reporting year. However, a complaint from 2018 about data access at the state statistical offices is still being processed: following a change in laws in 2017, the federal-level and state-level statistical offices can now make formally anonymised individual data available to researchers via guest researcher workstations. In Bavaria, this new opportunity was only partly used; important and highly sought-after data continued to be available only via controlled remote data processing – which tends to be inflexible and time-consuming for researchers.

The process was closed since the complaint was found to be unsubstantiated, seeing as the Bavarian data could indeed be accessed via controlled remote data processing. However, the RatSWD still regards the resulting restrictions as relevant to Germany as a research location, and has entered into (personal and written) contact with the Bavarian State Statistical Office and the Bavarian Ministry of the Interior. The process will continue within the RatSWD.

19 RatSWD [German Data Forum] (2018): The German Data Forum (RatSWD) and Research Data Infrastructure: Status Quo and Quality Management. RatSWD Output 1 (6). Berlin: German Data Forum (RatSWD). https://doi.org/10.17620/02671.30

20 https://www.konsortswd.de/en/datacentres/monitoring-and-complaints-management/complaints-office

Appendix

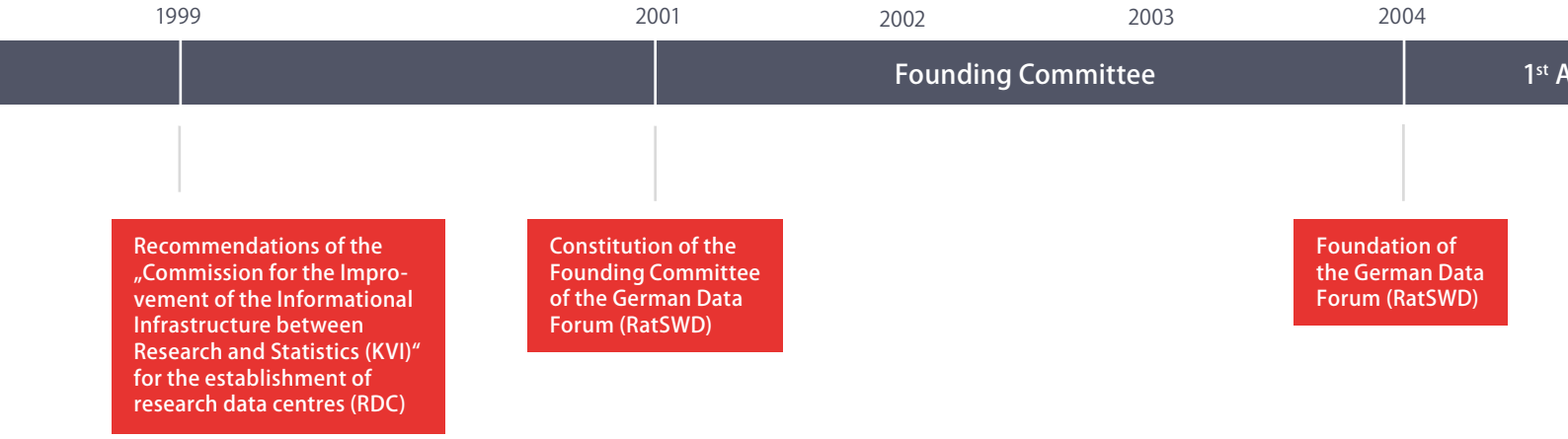
Appendix A

Development of the RatSWD's research data infrastructure and RDCs

Last update: 20/05/2020

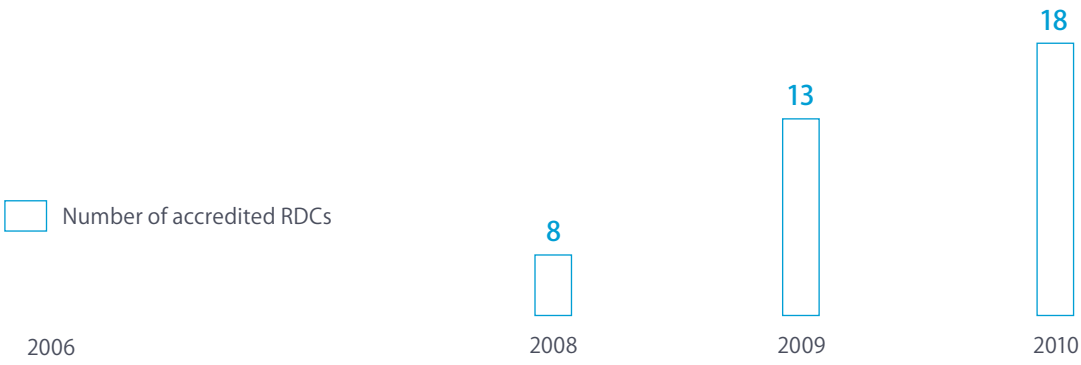
The research data centres Federal Statistical Office, Statistical Offices of the Länder, GML, IZA, BA at IAB, and RV were established prior to the foundation of the German Data Forum (RatSWD) and became part of the research data infrastructure in 2004. In these cases, the year of the RDCs' foundation is listed. All other RDCs were accredited after 2004 by the German Data Forum (RatSWD). With these RDCs, the year provided is the year of their accreditation.

The following RDCs are founded:



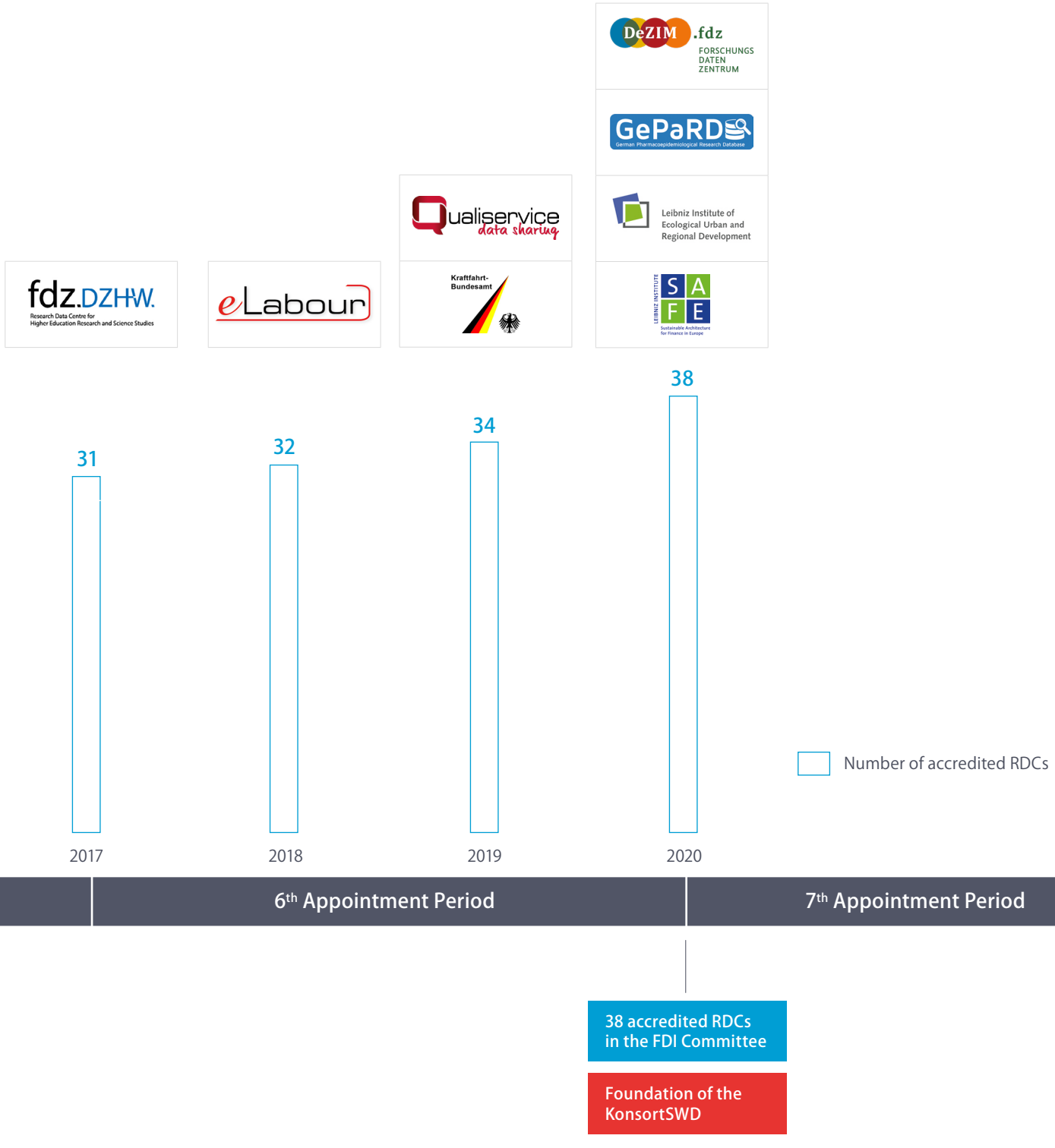
The following RDCs are accredited:

- SOEP The Socio-Economic Panel
- Datenservicezentrum Betriebs- und OrganisationsDATEN
- SHARE Survey of Health, Ageing and Retirement in Europe 50+ in Europe
- DZA German Centre of Gerontology
- @gesis fdz International Survey Programms
- leibniz-psychology.org
- IQB Institut zur Qualitätsentwicklung im Bildungswesen Forschungsdatenzentrum
- @gesis fdz Elections
- pairfam
- bibb Federal Institute for Vocational Education and Training Research Data Centre
- @gesis fdz ALLBUS
- fdz ruhr am rwi



ppointment Period	2nd Appointment Period	3rd Appointment Period
	8 accredited RDCs	Establishment of the Standing Committee on Research Data Infrastructure (FDI Committee) of the German Data Forum (RatSWD)
		Implementation of accreditation criteria and minimum standards

The following RDCs are accredited:


fdz.DZHW.
Research Data Centre for Higher Education Research and Science Studies
eLabour
Qualiservice data sharing
Kraftfahrt-Bundesamt
DeZIM .fdz
FORSCHUNGS DATEN ZENTRUM
GePaRD
German Pharmacoepidemiological Research Database
Leibniz Institute of Ecological Urban and Regional Development
LEIBNIZ INSTITUTE SAFE Sustainable Architecture for Finance in Europe
31
32
34
38
2017
2018
2019
2020
Number of accredited RDCs
6th Appointment Period
7th Appointment Period
38 accredited RDCs in the FDI Committee
Foundation of the KonsortSWD

Appendix B

Index and categories of data of the research data infrastructure oft he RatSWD

Last update: 21/07/2020

Available: Social Economic Education Health Behavioural Qualitative Other

BIBB-FDZ

Research Data Centre of the Federal Institute for Vocational Education and Training

Firm-level and individual-level datasets of vocational education research dealing primarily with the attainment and use of vocational knowledge and skills.

https://www.bibb.de/en/53.php

DeZIM.fdz *(provisional accreditation)*

Research Data Centre of the German Centre for Integration and Migration Research

The DeZIM.fdz organises access to research data collected at the German Centre for Integration and Migration Research. Additionally, the DeZIM.fdz offers comprehensive support on this data and on various methodological key issues.

https://dezim-institut.de/forschungsdatenzentrum-dezimfdz

EBDC

LMU-ifo Economics & Business Data Center

Datasets of German companies, including survey data collected by the ifo Institute on firms' business status, innovativeness, and investment behaviour, as well as external data on corporate financing and governance structure. Merged panels of the aforementioned two data sources are also available.

https://www.ifo.de/en/EBDC

FDZ AGD

Research Data Center Archive for Spoken German at the Institute for the German Language

Data on spoken German in interactions (conversation corpora) and data on domestic and non-domestic varieties of German (variation corpora).

http://agd.ids-mannheim.de/index_en.shtml

FDZ BA at IAB

Research Data Centre of the German Federal Employment Agency at the Institute for Employment Research

Data on persons, households, and employers, as well as combined datasets consisting of survey data and administrative research data in the fields of social security and labour market, and employment research.

https://fdz.iab.de/en.aspx

FDZ Bildung

Research Data Centre for Education at the DIPF | Leibniz Institute for Research and Information in Education

The hosted datasets include approaches of qualitative educational research such as video data, transcriptions, contextual materials from observations and interviews and survey tools of quantitative educational research such as questionnaires and assessment tests. The collected datasets refer mainly to the quality of instruction and to the quality of schools but also cover all levels of education throughout the entire span of life.

https://www.fdz-bildung.de/home?la=en

FDZ-BO

Research Data Centre for Business and Organizational Data

Quantitative and qualitative business, organizational data, linked employer and employee data, and data from employee and member surveys.

http://www.fdz-bo.diw.de

FDZ-Bund

Research Data Centre of the Federal Statistical Office

Germany-wide access to official statistics microdata from the following fields: population, education, health, business, agriculture, environment, administration of justice, finance, and taxes.

https://www.forschungsdatenzentrum.de/en

FDZ BZgA

Research Data Centre of the Federal Centre for Health Education

Data from nationally representative surveys, repeated at regular intervals, measuring the population's susceptibility to health education and prevention campaigns, as well as the knowledge, attitudes, and behaviour in the general population concerning the health issues addressed by BZgA.

https://www.bzga.de/home/bzga

FDZ-DJI

Research Data Centre of the German Youth Institute

Data from the surveys on children and young people growing up and the life situations of adults and families, conducted in regular intervals since 1988.

https://www.dji.de/abt2

FDZ-DZA

Research Data Centre of the German Centre of Gerontology

Data from the long-term German Ageing Survey (DEAS) on the changing life situations and ageing processes of people in mid- and older adulthood, and from the German Survey on Volunteering (FWS), a representative survey programme with a focus on voluntary activities and civic participation in Germany.

https://www.dza.de/en/fdz.html

fdz.DZHW

Research Data Centre for Higher Education Research and Science Studies

Quantitative and qualitative research data from the field of higher education and science studies, especially the DZHW Panel Study of School Leavers with a Higher Education Entrance Qualification (Studienberechtigtenpanel), the DZHW Graduate Panel (Absolventenpanel), the DZHW Social Survey, and the DZHW Science Survey.

https://www.fdz.dzhw.eu/en

FDZ eLabour

Research Data Centre eLabour

Qualitative data from studies in industrial and occupational sociology with a focus on the changing nature of work, including open and semi-standardised interviews, observations, and expert interviews.

http://elabour.de

FDZ GePaRD
(provisional accreditation)

German Pharmacoepidemiological Research Database

The FDZ GePaRD is based on data provided by statutory health insurance providers in Germany since 2004. GePaRD can be used to investigate research questions on the utilization and safety of drugs and vaccines in routine care, provided the respective data use has been approved in accordance with § 75 SGB X.

https://www.bips-institut.de/forschung/forschungsinfrastrukturen/gepard.html

Available: Social · Economic · Education · Health · Behavioural · Qualitative · Other

FDZ IQB

Research Data Centre of the Institute for Educational Quality Improvement

German datasets from the major national and international school performance studies and national studies measuring educational standards.

https://www.iqb.hu-berlin.de/fdz

FDZ IZA, IDSC

International Data Service Centre at the Institute for the Study of Labour

National and international labour market datasets with standardised information (eddi-conferences.eu). Research with, methods and resources for using online data for labor economics and social science. Development of tools and methods for remote access (statsdirect.org) and remote processing (JoSuA).

https://www.iza.org/en/research/idsc

FDZ at KBA

Research Data Centre at Kraftfahrt-Bundesamt

The research data centre provides anonymised quantitative microdata on additions to the Register of Driver Fitness (*Fahreignungsregister*) for scientific research.

https://www.kba.de/DE/Statistik/Forschungsdatenzentrum/forschungsdatenzentrum_node.html

FDZ-Länder

Research Data Centre of the Statistical Offices of the Länder

Germany-wide access to official statistics microdata from the following fields: population, education, health, business, agriculture, environment, administration of justice, finance, and taxes.

https://www.forschungsdatenzentrum.de/en

FDZ pairfam

Research Data Centre of the German Family Panel

Datasets from the "Panel Analysis of Intimate Relationships and Family Dynamics" (pairfam), a representative, interdisciplinary longitudinal study for the analysis of private living arrangements in Germany.

https://www.pairfam.de/en

FDZ PsychData at ZPID

Research Data Centre PsychData of the Leibniz Institute for Psychology Information

Pooled quantitative datasets from both basic research and applied psychology; data archiving with a focus on longitudinal studies, large-scale survey studies, and development testing.

https://www.psychdata.de/index.php?main=none&sub=none&lang=eng

FDZ Qualiservice

Research Data Centre Qualiservice

Qualiservice focuses on archiving, curating and providing qualitative research data from a range of disciplines. Its secure, flexible, and research-oriented services include processing primary qualitative studies for secondary use, comprehensive user support, long-term preservation, and the provision of archived research data as well as relevant context information.

https://www.qualiservice.org

FDZ Ruhr at RWI

Research Data Centre Ruhr at the RWI – Leibniz Institut for Economic Research

Specialisation on regional data: socioeconomic data measured by 1 square km grids. Aside from geo-referencing data on a scientific basis, the RDC provides various individual-level and employer-level data collected in RWI research projects.

https://en.rwi-essen.de/forschung-und-beratung/fdz-ruhr

FDZ-RV

Research Data Centre of the German Pension Insurance

Data on the insurance accounts of individuals insured in the Federal Pension Insurance. The accounts contain data on the insured persons' insurance history and the pension and rehabilitation benefits they received.

http://forschung.deutsche-rentenversicherung.de/FdzPortalWeb/dispcontent.do?id=main_fdz_english

FDZ SFB 882

(2016 dissolved)

Qualitative and quantitative datasets from inequality research

(The RDC was discontinued in 2016. Depending on the data type and basis, the data of RDC SFB 882 were transferred to different organisations: IAB data were handed over to FDZ BA im IAB; qualitative data with an organisational connection were handed over to FDZ BO; the remaining data were handed over to the SOBI archive at the University of Bielefeld (currently under development). (Last update on 09/14/2017)

https://sfb882.uni-bielefeld.de/en/fdz-sfb882.html

FDZ-SHARE

Research Data Centre of the Survey of Health, Ageing and Retirement in Europe

Data from the multidisciplinary and cross-national panel study "Survey of Health, Ageing and Retirement in Europe" (SHARE), which produces microdata on health, socio-economic conditions, and social and family networks of approximately 140,000 individuals in its seventh wave aged 50 or older in more than 20 European countries and Israel. The eighth wave of SHARE was collected in 2020.

http://www.share-project.org/data-access.html

IOER Monitor

(provisional accreditation)

Monitor of Settlement and Open Space Development

The IOER Monitor is a service of the Leibniz Institute for Ecological Urban and Regional Development (IOER). It provides data and information on the sustainability of land cover and land use change and for the landscape quality for the whole of Germany.

https://www.ioer-monitor.de/en

RDC ALLBUS

Research Data Centre ALLBUS at GESIS

Data from the Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) and German General Social Survey (GGSS) in English, on the attitudes, behaviours, and social structure of the German population.

https://www.gesis.org/en/allbus/allbus-home

RDC Elections

Research Data Centre Elections at GESIS

ZAccess to German national election surveys (federal elections and state elections), Politbarometer, Forsa-Bus, ARD Deutschlandtrend and Surveys for the Federal Government. The RDC's largest project at this point is the German Longitudinal Election Study (GLES).

https://www.gesis.org/en/elections-home

Available: Social · Economic · Education · Health · Behavioural · Qualitative · Other

RDC GML

Research Data Centre German Microdata Lab at GESIS

Research based services for researchers working with microdata from European and German official statistics: tools for data management and data analysis. Metadata (MISSY): comprehensive data documentation for official microdata on a detailed level. Knowledge transfer: consulting, training, workshops and user conferences on methodological and substantive research questions in the analysis of official microdata. Established 1987.

https://www.gesis.org/en/gml/gml-home

RDC International Survey Programmes

Research Data Centre International Survey Programmes at GESIS

Internationally comparative survey data from more than 70 countries on nearly all social science topics: Comparative Study of Electoral Systems (CSES), European Values Study (EVS), Eurobarometer, European Election Studies (EES), International Social Survey Programme (ISSP).

https://www.gesis.org/en/institute/research-data-centers/rdc-international-survey-programs

RDC-IWH

Research Data Centre of the Halle Institute for Economic Research

Company data from panel studies and longitudinal studies on development trends in the manufacturing and construction sectors of Eastern Germany, on privatisation activities of the *Treuhandanstalt*, on the choice of location for multinational companies in Eastern and Central Europe, and productivity and competitiveness indicators of European countries.

https://www.iwh-halle.de/en/research/data-and-analysis/research-data-centre

RDC-LIfBi

Research Data Center of the Leibniz Institute for Educational Trajectories at the University of Bamberg

Longitudinal data from the National Educational Panel Study (NEPS), which was launched in 2010 with more than 60,000 panel participants in six starting cohorts to study skills formation, educational processes, educational decisions, and educational returns in formal, non-formal, and informal contexts across the lifespan.

https://www.lifbi.de/Institute/Organization/Research-Data-Center

RDC PIAAC

Research Data Center Programme for the International Assessment of Adult Competencies (PIAAC) at GESIS

German and international data of the Programme for the Assessment of Adult Competencies (PIAAC). For Germany, additional regional data and longitudinal data are available.

https://www.gesis.org/en/piaac/rdc

RDC RKI

Research Data Centre of the Robert Koch Institute

Data on the state of health and health-related behaviour of Germany's resident population, collected on the basis of nationally representative studies.

https://www.rki.de/puf

RDC SOEP

Research Data Center of the Socio-Economic Panel Study at DIW Berlin

Data from representative annual surveys of private households. The SOEP-CORE sample features topics such as income, employment, education, and health. In addition, there is the longitudinal innovative sample (SOEP-IS), which enables external researchers to contribute research projects of their own.

https://www.diw.de/en/diw_02.c.222518.en

RDC Wissenschafts-statistik

Research Data Center Wissenschaftsstatistik of the Stifterverband

Data on the research and development activities of German companies, on the financial volume, structure, and regional distribution of research and development activities (R&D), and on R&D staff in the business sector.

https://www.fdz-wissenschaftsstatistik.de

RDSC Bundesbank

Deutsche Bundesbank Research Data and Service Centre

Various datasets on banks, securities, investment funds and enterprises, as well as combinations of those; panel survey on household finances.

https://www.bundesbank.de/en/bundesbank/research/rdsc

SAFE RDC
(provisional accreditation)

Research Data Center of the Leibniz Institute for Financial Research SAFE

The lack of pan-European financial data means that researchers have to resort to US data and cannot easily transfer research results to the European area. The SAFE Research Data Center not only pools existing data, but also collects and creates new German and European data sets to strengthen the European perspective of empirical research.

https://safe-frankfurt.de/data-center.html

ZEW-FDZ

ZEW Research Data Centre for European Economic Research

The ZEW-FDZ provides microdata from ZEW firm surveys on innovation activities, the development of young firms, the use of information and communication technologies, and further topics. Data from individual and expert surveys are also accessible – for example, the ZEW Financial Market Survey.

https://kooperationen.zew.de/en/zew-fdz/home.html

RDC WSI
(Guest status at FDI Committee)

Research Data Centre of the Hans-Böckler-Foundation at the WSI

The RDC WSI provides access to the data of the WSI Works Councils Survey until 2011.

https://www.wsi.de/en/index.htm

Appendix C

The monitoring commission

For quality assurance purposes, the German Data Forum (RatSWD) agreed to establish a monitoring commission in July 2016. Its main task is to collect and assess the regular reports handed in by the RDCs. Moreover, the commission monitors compliance with the obligations arising from provisional accreditation. The FDI Committee elects the commission from its own membership for a three-year term concurrent with the German Data Forum (RatSWD) appointment period. The commission thus enjoys a special level of trust and legitimacy. It consists of four members of the FDI Committee and two deputy members (to replace elected members, if required) and the German Data Forum (RatSWD) chairpersons sit in as guests.

Members of the monitoring commission

Maurice Brandt (Chair: May 2019 – 31 December 2019)
Research Data Centre of the Federal Statistical Office (FDZ-Bund)

Dr. Lea Eilers (Chair: July 2018 – September 2018)
Research Data Centre Ruhr at the RWI – Leibniz Institute for Economic Research (FDZ Ruhr at RWI)

Dr. Cornelia Lang (Chair: since 1 January 2020)
Halle Institute for Economic Research (RDC-IWH)

Holger Quellenberg (Chair: October 2018 – April 2019)
Research Data Centre of the German Youth Institute (FDZ-DJI)

Dr. Sandra Gottschalk (Deputy)
ZEW Research Data Centre for European Economic Research (ZEW-FDZ)

Tatjana Mika (Deputy)
Research Data Centre of the German Pension Insurance (FDZ-RV)

Standing guests of the monitoring commission

Prof. Stefan Bender
Vice chair of the German Data Forum (RatSWD) (July 2014 – November 2019)

Prof. Regina T. Riphahn, Ph.D.
Chair of the German Data Forum (RatSWD) (July 2014 – July 2020)

Prof. Dr. Jürgen Schupp
Vice chair of the German Data Forum (RatSWD) (November 2019 – July 2020)

Prof. Dr. Monika Jungbauer-Gans
Chair of the German Data Forum (RatSWD) (since July 2020)

Prof. Dr. Kerstin Schneider
Vice chair of the German Data Forum (RatSWD) (since July 2020)

Appendix D

Contributors to the 2019 Activities Report

Maurice Brandt (Member of the monitoring commission)
Research Data Centre of the Federal Statistical Office (FDZ-Bund)

Dr. Mathias Bug (RatSWD business office)

Dr. Lea Eilers (Member of the monitoring commission)
Forschungsdatenzentrum Ruhr am RWI – Leibniz-Institut für Wirtschaftsforschung (FDZ Ruhr am RWI)

Dr. Anna Fräßdorf (RatSWD business office)

Dr. Sandra Gottschalk (Deputy member of the monitoring commission)
ZEW Research Data Centre (ZEW-FDZ)

Dr. Cornelia Lang (Chair of the monitoring commission)
Research Data Centre of the Halle Institute for Economic Research (RDC-IWH)

Tatjana Mika (Deputy member of the monitoring commission)
Research Data Centre of the German Pension Insurance (FDZ-RV)

Holger Quellenberg (Member of the monitoring commission)
Research Data Centre of the German Youth Institute (FDZ-DJI)

Dr. Katrin Schaar (RatSWD business office)

Imprint

Publisher:
German Data Forum (RatSWD)
Rungestr. 9
10179 Berlin
Germany
office@ratswd.de
https://www.ratswd.de

Editors:
Dr. Katrin Schaar, Dr. Mathias Bug

Layout:
Claudia Kreutz

Translation:
Simon Wolff | translate.berlin

Icons:
made by Freepik from https://www.flaticon.com

Berlin, October 2020

RatSWD Output:
The RatSWD Output Series documents the German Data Forum's (RatSWD) activities during its 6th appointment period (2017–2020). It serves to publish its statements and recommendations and to make them available to a broad audience.

This report is the result of a project that is funded by the Federal Ministry for Education and Research (reference number: 01UW1802). Unless otherwise stated, the responsibility for this publication lies with the German Data Forum (RatSWD).

doi: 10.17620/02671.59

Citation suggestion:
RatSWD [German Data Forum] (2020): Activities Report 2019 of the Research Data Centres (RDCs) accredited by the German Data Forum (RatSWD). Berlin, German Data Forum (RatSWD). https://doi.org/10.17620/02671.59

Established in 2004, the **German Data Forum** (Rat für Sozial- und Wirtschaftsdaten, RatSWD) is an independent council. It advises the German federal government and the federal states (Länder) in matters concerning the research data infrastructure for the empirical social, behavioural, and economic sciences. The German Data Forum (RatSWD) has 16 members. Membership consists of eight elected representatives of the social, behavioural, and economic sciences and eight appointed representatives of Germany's most important data producers.

The German Data Forum (RatSWD) offers a forum for dialogue between researchers and data producers, who jointly issue recommendations and position papers. The council furthers the development of a research infrastructure that provides researchers with flexible and secure access to a broad range of data. The German Data Forum (RatSWD) has accredited 38 research data centres (as of May 2020) and fosters their interaction and collaboration.

www.ratswd.de